SLEEPJ, 2020, 1–11

doi: 10.1093/sleep/zsaa090
Advance Access Publication Date: 10 July 2020
Original Article

ORIGINAL ARTICLE

Optimizing actigraphic estimates of polysomnographic sleep features in insomnia disorder

Bart H. W. te Lindert[1,*], Wisse P. van der Meijden[1], Rick Wassing[1,2], Oti Lakbila-Kamal[1], Yishul Wei[1], Eus J. W. Van Someren[1,3,4,†], Jennifer R. Ramautar[1,†]

[1]Department of Sleep and Cognition, Netherlands Institute for Neuroscience, an institute of the Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands, [2]Sleep and Circadian Research Group, Woolcock Institute of Medical Research, University of Sydney, Camperdown, NSW, Australia, [3]Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, The Netherlands and [4]Department of Psychiatry, Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

*Corresponding author. Bart H. W. te Lindert, Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands. Email: b.te.lindert@gmail.com

†These authors contributed equally to this work.

Abstract

Study Objectives: Actigraphy is a useful tool for estimating sleep, but less accurately distinguishes sleep and wakefulness in patients with insomnia disorder (ID) than in good sleepers. Specific algorithm parameter settings have been suggested to improve the accuracy of actigraphic estimates of sleep onset or nocturnal sleep and wakefulness in ID. However, a direct comparison of how different algorithm parameter settings affect actigraphic estimates of sleep features has been lacking. This study aimed to define the optimal algorithm parameter settings for actigraphic estimates of polysomnographic sleep features in people suffering from ID and matched good sleepers.

Methods: We simultaneously recorded actigraphy and polysomnography without sleep diaries during 210 laboratory nights of people with ID (n = 58) and matched controls (CTRL) without sleep complaints (n = 56). We analyzed cross-validation errors using 150 algorithm parameter configurations and Bland–Altman plots of sleep features using the optimal settings.

Results: Optimal sleep onset latency and total sleep time (TST) errors were lower in CTRL (8.9 ± 2.1 and 16.5 ± 2.1 min, respectively) than in ID (11.7 ± 0.8 and 29.1 ± 3.4 min). The sleep–wake algorithm, a period duration of 5 min, and a wake sensitivity threshold of 40 achieved optimal results in ID and near-optimal results in CTRL. Bland–Altman plots were nearly identical for ID and controls for all common all-night sleep features except for TST.

Conclusion: This systematic evaluation shows that actigraphic sleep feature estimation can be improved by using uncommon parameter settings. One specific parameter setting provides (near-)optimal estimation of sleep onset and nocturnal sleep across ID and controls.

Statement of Significance

Actigraphy distinguishes sleep and wakefulness less well in insomnia disorder (ID) than in good sleepers. Evaluation of a wide range of parameter settings revealed that actigraphic estimates can be improved. Optimal settings differed between people with ID and good sleepers. With minimal loss of accuracy, however, a single configuration can be recommended for use across good sleepers and people with ID.

Key words: actigraphy; insomnia; validation; polysomnography; sensitivity

Submitted: 2 May, 2019; **Revised:** 17 November, 2019

Introduction

Insomnia disorder (ID) is the second most common mental disorder in Europe [1] and is characterized by subjective reporting of difficulty initiating/maintaining sleep or early morning awakening despite adequate opportunity for sleep and resultant in daytime impairment [2, 3]. Objective estimates of sleep features are not required for the diagnosis of ID per se, but are considered valuable in the clinic when patients do not respond to cognitive behavioral therapy for insomnia, or when sleep features are required over prolonged periods [4]. In such cases, actigraphy could provide a more feasible and cost-effective method than the gold-standard polysomnography (PSG) to assess sleep features [4].

Actigraphy is the continuous recording of movement, usually of the wrist. Actigraphy can be used for many purposes like estimating physical activity, diurnal activity rhythms, severity of disordered, movement, and sleep. Depending on the purpose, valid application requires optimization of the recording and processing of the movement signal [5, 6]. We here address the important question of how to best use actigraphy to estimate sleep in ID.

The application of actigraphy in sleep research is based on the fact that prolonged periods of immobility are more likely to occur during sleep than during wakefulness. Actigraphy discriminates wakefulness from sleep based on the movement of a limb, but cannot discriminate between different sleep stages. Actigraphy therefore only allows for the calculation of all-night sleep features such as total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE).

A recent meta-analysis of the literature on the validity of actigraphy in ID [7] showed that actigraphic estimates were not always reliable enough and, for example, failed to detect intervention effects on sleep that were established by PSG. Of the aforementioned four common all-night common sleep features, only TST and SOL were considered valid features for use in clinical care decisions [7]. The inconsistent validity of actigraphic estimates may be explained by several factors including the technical specification of the device, the location of the device, the acquisition settings, clinical features of the population being studied, procedures for setting rest interval and data acquisition, and the algorithm used.

The actigraphy algorithm allows for flexible parameter tuning to improve estimation of sleep onset and sleep–wake discrimination. Sleep–wake discrimination can be tuned by changing the wake sensitivity threshold (WST) above which recorded activity is scored as wake. Sleep onset estimation can be tuned by changing the duration of consecutive immobility or estimated sleep required to define sleep onset. The optimal setting for sleep–wake discrimination may not necessarily be the optimal setting for sleep onset estimation and may moreover differ between populations.

Default parameters are frequently used to obtain actigraphic estimates sleep features in ID [8–10]. Several studies have evaluated parameter settings other than the default configurations. In line with the principle that patients with ID may lie still in bed while being awake, a low WST showed better concordance with PSG in elderly women with ID [11]. One study suggested that a low WST resulted in the best concordance between actigraphy and PSG sleep features, but no data were presented to support the hypothesis [12]. In chronic obstructive pulmonary disease (COPD) patients with comorbid ID, the default immobility period duration (PD) of 10 min with a very low WST was found to best match simultaneously recorded PSG features [13]. In a large sample of people with ID and good sleepers, actigraphic estimates of sleep features obtained with a low WST best discriminated ID from good sleepers, but no simultaneous PSG was recorded [14]. In contrast, a high WST was found to be most sensitive in young adults with ID [15].

In most studies, the required duration of immobility to mark sleep onset was either not reported [9, 12, 14] or was set to the default of 10 min [11, 15]. Some studies evaluated how reliability of the actigraphic estimate of sleep onset changed across a range of immobility PDs. As compared to 5 and 15 min, a PD of 10 min resulted in optimal sleep onset estimation in COPD patients with ID [13]. In other populations, the influence of immobility PD on sleep onset has been evaluated using several shorter and longer durations (3–20 min). A shorter PD of 5 min resulted in the best sleep onset estimation in patients with adult obstructive sleep apnea (OSA) [16], while medium to long PDs (10–20 min) were optimal in children and adolescents with or without OSA [17, 18].

To the best of our knowledge, no study has evaluated a wide range of algorithm parameters to optimize sleep onset estimation and sleep–wake discrimination in ID and controls using simultaneously recorded PSG and actigraphy. Therefore, the aim of the study was to find the optimal parameters to derive estimates of sleep features from actigraphy as compared to PSG in ID and matched good sleepers using a wide range of algorithm parameter settings while keeping all other factors constant.

Methods

Participants

To evaluate the agreement of actigraphic estimates of common sleep features with their gold-standard polysomnographic assessment, we recorded actigraphy and PSG simultaneously during several studies in our sleep laboratory [19, 20]. The studies were approved by the ethics committee of the VU University Medical Center, Amsterdam, The Netherlands. Participants were recruited through advertisement and the Netherlands Sleep Registry (https://www.sleepregistry.org) [21]. Screening was performed using online questionnaires, telephone, and a structured interview, including the Insomnia Severity Index (ISI) [22]. All participants provided written informed consent. The inclusion criteria for the ID group (n = 58, age range 21–69 years) conformed to the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition [2], *International Classification of Sleep Disorders*, 3rd Edition [23], and Research Diagnostic Criteria for Insomnia Disorder [24] as well as an ISI score of at least 10. The controls (CTRL) included age- and sex-matched volunteers (n = 56, age range 22–70 years) that reported to have no sleep difficulties, further confirmed by an ISI score less than 10. Exclusion criteria for all participants were: (1) diagnosed sleep apnea, restless legs syndrome, narcolepsy, or other somatic, neurological, or psychiatric disorders; (2) use of sleep medications within the prior 2 months including the recording days. Participants showing signs of sleep apnea or restless legs during laboratory PSG assessments were not included in the selected 114 participants. Table 1 summarizes the demographic characteristics and ISI scores of the sample.

Table 1. Demographics and polysomnographic sleep features (mean ± standard deviation)

Characteristic	Control (n = 56)	ID (n = 58)	P
Sex, female/male	39/17	44/14	.53
Age, years	43.2 ± 15.0	47.8 ± 14.0	.087
ISI	**2.5 ± 2.5**	**16.8 ± 3.8**	**<.0001**
TIB, min	482.1 ± 47.6	473.2 ± 66.1	.29
SOL, min	**12.1 ± 15.9**	**16.9 ± 17.1**	**.008**
TST, min	**429.0 ± 47.7**	**392.8 ± 83.3**	**.0004**
WASO, min	**32.0 ± 24.9**	**52.4 ± 47.3**	**.002**
SE, %	**89.1 ± 6.8**	**82.4 ± 14.6**	**.0003**

Group differences were tested with Fisher exact test for sex and by Wilcoxon rank-sum tests for the other variables. Bold font highlights significant group differences.
ISI, Insomnia Severity Index; TIB, time in bed; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; SE, sleep efficiency.

Protocol

People with ID and matched CTRL completed two consecutive nights of PSG and actigraphy in a laboratory setting. In total, 18 of the participants (10 with ID and 8 CTRL) had missing PSG or actigraphy data for one night, resultant in 9% missing data and a total of 210 individual overnight paired observations. On the recording days, participants were asked to refrain from alcohol and drugs, as well as to limit consumption of caffeinated beverages to a maximum of two cups, which were allowed only before noon. The lights-off and lights-on times for each participant were self-chosen according to individual habitual bedtime and did not significantly differ between the two groups (mean ± standard deviation: ID = 23:23 ± 00:42, CTRL = 23:27 ± 00:41 h, p = .61 and ID = 07:16 ± 01:03, CTRL = 07:29 ± 00:42 h, p = .16, respectively).

Polysomnography

PSG was recorded in each participant using a 256-channel LTM HydroCel Geodesic Sensor Net and a Polygraphic Input Box (Electrical Geodesics Inc., Eugene, Oregon) connected to a Net Amps 300 amplifier (input impedance: 200 MΩ, A/D converter: 24 bits). All PSG recordings were visually scored offline by experienced scorers (J.R.R. and O.L.K.) blind to the participants' group classification. Inter-scorer agreement in our lab generally lies between 0.67 and 0.80 (mean = 0.72 and 0.76 for ID and CTRL, respectively) [19]. Scoring of sleep stages was based on signals obtained from six electroencephalogram leads (electrode # 36, 224, 72, 173, 116, 150 of the HydroCel Geodesic Sensor Net, approximately equivalent to F3, F4, C3, C4, O1, O2, respectively, in the 10–20 system) and two electro-oculogram leads (1 cm below the left and above the right outer canthi) referenced to linked mastoids and one bipolar chin electromyogram channel. Each 30-s epoch was scored as wakefulness (W), stage 1 sleep (N1), stage 2 sleep (N2), stage 3 sleep (N3), or REM sleep (R), according to the American Academy of Sleep Medicine (AASM) manual [25]. The PSG recording was started at lights off and stopped at lights on. To calculate epoch-by-epoch agreement of sleep and wakefulness between PSG and actigraphy, epochs scored as N1, N2, N3, or R were marked as sleep and epochs scored as W were marked as wake.

Actigraphy

Actigraphy was recorded using a microelectromechanical system (MEMS) accelerometer (GENEActiv Sleep, Activinsights Ltd., Kimbolton, UK) at a sampling frequency of 50 Hz. The GENEActiv recorder contains a tri-axial MEMS accelerometer with a measurement range of ±8 g and a sensitivity of ≥0.004 g and records motion-related and gravitational acceleration. Data were uploaded to a PC using the GENEActiv PC software (version 1.0, Activinsights Ltd.).

Temporal alignment

Optimal alignment of PSG and actigraphy was obtained by cross-correlating the binary coded sleep/wake signal from PSG with the binary coded immobile/mobile actigraphy signals.

Algorithm parameter settings

Sleep–wake discrimination

So-called "activity counts" were obtained using a validated method [26] adapted to 30-s epochs. Subsequently, each epoch was classified as either sleep or wake. The activity in each epoch was re-scored by weighing activity in the surrounding 2-min period. For each 30-s epoch, the activity was weighted as follows [27, 28]:

$$A_0 = 0.04E_{-(4-3)} + 0.2E_{-(2-1)} + 4E_0 + 0.2E_{+(1-2)} + 0.04E_{+(3-4)},$$

where A_0 is the total re-scored activity for the 30-s epoch of interest; E_0 is the activity in the scored epoch; E_n is the activity in the four epochs before (E_{-4} to E_{-1}) and after (E_{+1} to E_{+4}) the scored epoch. If A_0 is less than or equal to a predefined WST ($A_0 \leq WST$) the epoch is scored as sleep, otherwise ($A_0 > WST$) the epoch is scored as wake (Table 2). Sleep–wake discrimination was evaluated across a range of WSTs (10–100 in intervals of 10, where 20, 40, and 80 are commonly denoted as the default low, medium, and high WSTs, respectively [28]).

Epoch-by-epoch agreement

Actigraphy epochs were classified as true sleep (TS), false sleep (FS), true wake (TW), and false wake (FW) based on their agreement with the corresponding PSG epochs (Table 3). Based on these classification, the accuracy, sensitivity, specificity, positive predictive value (PPV or precision), negative predictive value (NPV), and Cohen's kappa statistics (Table 4) were calculated at each WST to quantify epoch-by-epoch agreement [29, 30]. Agreement, sensitivity, and specificity indicate the proportions of all epochs correctly classified as wake or sleep by actigraphy compared with PSG. PPV and NPV indicate the proportion of sleep or wake epochs, respectively, correctly classified by actigraphy. Cohen's kappa (κ) indicates agreement beyond what can be expected by chance alone.

Sleep onset estimation

Sleep onset defines the beginning of the assumed sleep period (ASP). Sleep onset can be either not estimated (*none*) or estimated by choosing one of two different methods: *immobile–mobile* and *sleep–wake* (Table 2). The *immobile–mobile* algorithm as implemented in the Actiware software searches for the first immobility PD of *X* min after bedtime with no more than one epoch containing *any* activity count. The outcome only depends on the chosen parameter *X*. In contrast, the *sleep–wake* algorithm searches for the

Table 2. Algorithm parameters

Setting	Value	Description
WST	10–100 at intervals of 10	The threshold above which activity is scored as wake
PD	3, 5, 10, 15, or 20 min	The duration of the period in minutes used to estimate sleep onset. Only applicable if the immobile–mobile or sleep–wake algorithm is used to estimate sleep onset
Sleep onset estimation algorithm	Immobile–mobile algorithm	Calculates sleep onset based on the presence of movement or no movement, independently of the sleep–wake discrimination. Sleep onset was defined using the first immobile period after bedtime of at least *X* min with no more than one epoch containing any movement. The first epoch of this period was classified as sleep onset. Therefore, this sleep parameter was not influenced by the WST
	Sleep–wake algorithm	Calculates sleep onset based on sleep–wake discrimination. Sleep onset was defined using the first sleep period after bedtime of at least *X* min with continuous sleep. The first epoch of this period was classified as sleep onset. Therefore, this sleep parameter is influenced by the WST
	No sleep onset estimation	Sleep onset is not estimated and equal to bedtime

Table 3. Confusion matrix used in the calculation of agreement measures

	PSG	
Actigraphy	Sleep	Wake
Sleep	True sleep (TS)	False sleep (FS)
Wake	False wake (FW)	True wake (TW)

first PD of *X* min of continuous estimated sleep after bedtime. The outcome therefore depends on both the chosen parameter *X* and the WST chosen to define wakefulness. Ignoring sleep onset estimation (*none*) means the ASP is equal to the period between lights off and lights on. To evaluate the effects of different parameter settings, sleep onset was estimated based on PDs of 3, 5, 10, 15, and 20 min at 10 equidistant WSTs between 10 and 100.

Sleep feature calculation

Estimates of the most commonly used sleep features (Table 5)—SOL, TST, WASO, and SE—were calculated using 150 different configurations with 10 WSTs (10–100 at intervals of 10), 5 PDs (3, 5, 10, 15, and 20 min), and 3 sleep onset algorithms (*sleep–wake*, *immobile–mobile*, and *none*). Analyses of sleep features were performed using custom scripts (https://github.com/btlindert/actant-1) written in MATLAB 8.3 (The Mathworks Inc., Natick, MA).

Algorithm parameter tuning

Algorithm parameter tuning was performed separately for ID and CTRL. From each of the separate ID and CTRL datasets, 20% of the data were randomly selected and set apart as a holdout sample for final testing by splitting at the participant level. In the remaining data, still separately for ID and CTRL, fivefold cross-validation was used to find the algorithm parameter settings that resulted in the best agreement between PSG and actigraphy. In each of the fivefolds split at the participant level, a grid search on 80% of the training data defined the algorithm parameter combination that resulted in the closest match between actigraphic and PSG sleep feature estimates. These settings were then applied to the remaining 20% validation fold to obtain the mean absolute error (MAE) between PSG and actigraphic SOL or TST estimates. The MAE was calculated as:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\mathrm{PSG}_i - \mathrm{Actigraphy}_i|}{n},$$

where MAE is the absolute differences in sleep features (SOL or TST) obtained from PSG (*PSG*) and actigraphy (*Actigraphy*) for every night (*i*) averaged over all nights (*n*) of simultaneously recorded PSG and actigraphy.

The mean and standard error of the MAE were calculated across the fivefolds and will be denoted as "validation MAE." The configuration that was optimal over the fivefolds was used to evaluate performance in the 20% holdout test sample for final validation and the resulting MAE will be denoted as "test MAE." The influence of missing data was minimized by randomly distributing the participants with missing data across the different folds.

Optimal parameter set across ID and controls

In situations where the diagnosis of ID is not yet known or changes over time, a single configuration that achieves optimal results across ID and CTRL is valuable. We evaluated if the results allowed for a common optimal parameter set to be defined across ID and CTRL. The optimal parameter set for ID and CTRL was applied to all measurements to calculate all-night sleep feature agreement.

All-night sleep feature agreement

The agreement between all-night PSG-derived sleep features and the corresponding actigraphic estimates obtained using the optimal parameters across ID and CTRL was visually inspected with Bland–Altman (mean-difference) plots [31]. For each plot, an ordinary least squares (OLS) regression was used to test and correct for proportional or constant bias throughout the measurement range [32]. Heteroskedasticity of differences over the measurement range was evaluated with a Breusch–Pagan test of the residuals. If the variance of the differences was proportional to the mean, the limits of agreement (LOAs) were modified using the slope from an OLS regression of absolute residuals multiplied by $1.96^{*}\sqrt{(\pi/2)}$ [32].

Statistical analysis

Data preprocessing and calculation of agreement statistics and all-night sleep parameters were performed offline using custom-written MATLAB programs. All the other statistical analyses were conducted using R [33].

Table 4. Formulas and description of the agreement measures

Measure	Formula	Description
Accuracy	$[TS + TW/(TS + TW + FS + FW)] \times 100$	Percentage of all epochs correctly detected as wake or sleep by actigraphy.
Sensitivity	$[TS/(TS + TW + FS + FW)] \times 100$	Percentage of all epochs correctly detected as sleep by actigraphy.
Specificity	$[TW/(TS + TW + FS + FW)] \times 100$	Percentage of all epochs correctly detected as wake by actigraphy.
PPV	$[TS/(TS + FS)] \times 100$	Percentage of PSG sleep epochs correctly detected as sleep by actigraphy.
NPV	$[TW/(TW + FW)] \times 100$	Percentage of PSG wake epochs correctly detected as wake by actigraphy.

TS, true sleep; TW, true wake; FS, false sleep; FW, false wake; PPV, positive predictive value; NPV, negative predictive value; PSG, polysomnography.

Table 5. Description of sleep parameters

Setting	Description
Time in bed (TIB)	Time between bedtime and get-up time
Sleep onset	The time associated with the onset of sleep using either the *immobile–mobile* or *sleep–wake* algorithm
ASP	Time between sleep onset and final wake time
SOL	Time between lights off and sleep onset
TST	The number of epochs within the ASP scored as sleep multiplied by the epoch length
WASO	The number of epochs within the ASP scored as wake multiplied by the epoch length
SE	The ratio of TST to ASP multiplied by 100

Results

Epoch-by-epoch agreement

We calculated epoch-by-epoch agreement at all WSTs across the PSG recording (Table 6). Overall, a higher WST increased the accuracy, sensitivity and NPV but attenuated the specificity and PPV. These results are partially explained by the large proportion of sleep epochs (i.e. the likelihood that an epoch is sleep is much higher than the likelihood that it is wake) [29]. Cohen's kappa accounts for this by calculating the agreement beyond what can be expected by chance. Both for ID and CTRL, a WST of 20 provided the highest Cohen's kappa (0.47 ± 0.15 and 0.44 ± 0.15, respectively), for which the agreement is considered moderate [34].

Optimal parameters to estimate SOL

Actigraphic estimates of SOL and TST were calculated using 150 different configurations with 10 WSTs (10–100 at intervals of 10), 5 PDs (3, 5, 10, 15, and 20 min), and 3 sleep onset algorithms (*sleep–wake*, *immobile–mobile*, and *none*) across 210 complete overnight observations of simultaneous PSG and actigraphy. For each configuration, the test MAE between actigraphy and PSG was calculated for SOL and TST, separately for ID and CTRL.

For both ID and CTRL, SOL estimation was generally better at high WSTs and short-to-medium PDs irrespective of the algorithm used (Figure 1). The mean validation MAE was lower in CTRL than in ID for each combination of parameters, indicating better actigraphic estimation of SOL in CTRL than in ID. In contrast, the standard error was lower in ID than in CTRL, indicating that more reliable differences in SOL between actigraphic estimation and PSG were obtained in ID.

In CTRL, the best estimation of SOL resulted in a validation MAE of 8.9 ± 2.1 min (mean ± standard error) and was achieved using the *immobile–mobile* algorithm with a PD of 10 min, confirming the validity of the default parameters in the Actiware software (Figure 1, A). The results were confirmed in the test set with a test MAE of 11.6 min, in line with the fact that validation error was trained on more data and was likely to underestimate the test error.

In ID, the best estimation of SOL resulted in a validation MAE of 11.7 ± 0.8 min (mean ± standard error) and was achieved using the *sleep–wake* algorithm with a PD of 5 min and a WST of 40 (Figure 1, B), which is a 18% improvement compared to the default parameters. The test MAE was 11.2 min. For both CTRL and ID, the magnitude of the test MAEs were relatively large compared to the median SOL derived from PSG (9 and 13 min, respectively). Interestingly, the *immobile–mobile* algorithm, which is the frequently used default setting in the Actiware software, resulted in higher validation MAEs in ID, suggesting common use of suboptimal settings for the actigraphic estimation of sleep onset in ID.

Optimal parameters to estimate TST

TST estimation was better in CTRL than in ID irrespective of the type of algorithm used (Figure 2). In ID, the *sleep–wake* algorithm, a PD of 10 min, and a WST of 40 achieved the lowest validation MAE of 29.1 ± 3.4 min (mean ± standard error; Figure 2, A). In CTRL, the optimal configuration to estimate TST was the *sleep–wake* algorithm, a PD of 20 min, and a WST of 90 for CTRL with a validation MAE of 16.5 ± 2.1 (mean ± standard error; Figure 2, B). The test MAEs were 26.7 and 29.9 min for ID and CTRL, respectively. The optimal parameters are higher than usually reported in the literature. For CTRL, TST estimation would benefit from a higher WST and a longer PD, compared to the commonly used default WST of 40 and PD of 10 min.

Figure 2, C shows that if reliable estimates of lights off and lights on can be reliably obtained (e.g. from video) and TST is the sole variable of interest, negligible improvement in TST estimation is obtained by estimating sleep onset.

Optimal parameter set across ID and controls

In situations where the diagnosis of ID is not yet known or changes over time, it would be valuable to have a single configuration that achieves optimal results across ID and CTRL. As shown above, actigraphic estimates have lower precision in ID than in CTRL, and the loss of precision as a result of suboptimal parameter selection is greater in ID than in CTRL. Therefore, a single optimal configuration would consist of the optimal parameters for ID: the *sleep–wake* algorithm, a PD of 5 min, and a WST of 40. These parameters achieve optimal results for SOL and near-optimal results for TST in ID. These parameters also

Table 6. Mean and standard deviation of epoch-by-epoch agreement between actigraphy and PSG at all WSTs for ID and CTRL

	Accuracy (%)		Sensitivity (%)		Specificity (%)		PPV (%)		NPV (%)		Cohen's kappa	
WST	CTRL	ID	CTRL	ID	CTRL	ID	CTRL	ID	CTRL	ID	CTRL	ID
10	86.2 ± 5.7	84.3 ± 5.9	88.5 ± 6.3	87.9 ± 5.9	**67.8 ± 17.0**	**68.5 ± 19.3**	**95.5 ± 4.1**	**92.7 ± 8.3**	40.6 ± 18.2	49.1 ± 20.5	0.40 ± 0.15	0.44 ± 0.15
20	89.1 ± 4.7	86.5 ± 6.3	92.7 ± 4.5	92.0 ± 4.6	60.0 ± 18.6	61.2 ± 20.3	94.9 ± 4.5	91.5 ± 8.9	47.9 ± 19.6	55.9 ± 20.6	**0.44 ± 0.15**	**0.47 ± 0.15**
30	90.1 ± 4.5	87.2 ± 6.8	94.7 ± 3.5	94.0 ± 3.7	52.9 ± 19.4	55.3 ± 20.4	94.2 ± 4.9	90.7 ± 9.3	51.9 ± 20.7	59.9 ± 20.3	0.44 ± 0.16	0.46 ± 0.15
40	90.6 ± 4.5	87.5 ± 7.1	95.8 ± 2.9	95.3 ± 3.0	48.3 ± 19.5	50.8 ± 19.6	93.8 ± 5.0	90.0 ± 9.5	54.8 ± 21.3	63.0 ± 20.1	0.43 ± 0.16	0.46 ± 0.15
50	90.9 ± 4.6	**87.6 ± 7.5**	96.4 ± 2.6	96.1 ± 2.5	44.8 ± 19.0	46.8 ± 19.4	93.5 ± 5.2	89.4 ± 9.7	57.2 ± 21.7	65.1 ± 20.2	0.42 ± 0.16	0.44 ± 0.15
60	91.0 ± 4.7	87.5 ± 7.8	97.0 ± 2.3	96.7 ± 2.3	41.6 ± 18.6	43.3 ± 18.7	93.1 ± 5.3	88.9 ± 9.9	59.3 ± 22.4	66.9 ± 19.9	0.41 ± 0.16	0.43 ± 0.15
70	**91.1 ± 4.9**	87.4 ± 8.0	97.4 ± 2.1	97.1 ± 2.0	39.4 ± 18.2	40.8 ± 18.2	92.9 ± 5.5	88.6 ± 10.0	61.2 ± 22.9	68.6 ± 20.2	0.40 ± 0.16	0.41 ± 0.15
80	91.1 ± 5.0	87.3 ± 8.3	97.7 ± 1.9	97.4 ± 1.9	36.7 ± 17.5	38.3 ± 17.6	92.7 ± 5.6	88.2 ± 10.2	62.8 ± 23.0	69.4 ± 20.1	0.39 ± 0.16	0.40 ± 0.16
90	91.1 ± 5.1	87.1 ± 8.5	97.9 ± 1.8	97.7 ± 1.7	34.6 ± 17.5	36.2 ± 17.2	92.5 ± 5.7	87.9 ± 10.3	63.3 ± 24.1	70.3 ± 19.8	0.38 ± 0.16	0.39 ± 0.15
100	91.1 ± 5.2	87.1 ± 8.7	**98.1 ± 1.6**	**97.9 ± 1.6**	32.9 ± 16.9	34.6 ± 16.8	92.3 ± 5.7	87.7 ± 10.5	**64.7 ± 23.8**	**71.7 ± 20.3**	0.37 ± 0.16	0.38 ± 0.15

ID, insomnia disorder; CTRL, control; WST, wake sensitivity threshold; PPV, positive predictive value; NPV, negative predictive value.
Bold font highlights the highest agreement obtained for each of the measures.

achieve near-optimal results in CTRL, while the absolute errors are still substantially smaller than those obtained in ID.

Bland–Altman plots

Using the single optimal set of parameters for both ID and CTRL just mentioned, we visually inspected the agreement of SOL, TST, WASO, and SE between actigraphy and PSG using Bland–Altman plots (Figure 3). The bias of actigraphic estimation of SOL is negligible for short SOL. However, actigraphy increasingly underestimates longer SOLs, by even more than half an hour at a SOL of an hour. In addition to the larger bias at increasing SOL, the estimates also become less reliable as indicated by the widening of the LOAs. Similar to SOL, WASO bias is near zero for very small WASO, but actigraphy underestimates WASO by as much as 100 min at a WASO of 200 min. Consequently, actigraphy only slightly overestimates SEs above 90% but strongly overestimates lower SEs, by as much as 20% at a SE of 60%. The LOAs strongly increase with lower SEs.

In CTRL, the bias of actigraphic estimation of TST is negligible and the LOA is constant across the entire measurement range. In ID, actigraphy underestimates TST below a TST of 425 min with an average bias of 50 min at 250 min, but instead overestimates TST above 425 min with an average bias of 25 min at 550 min. The reliability of the estimate improves at higher TST, as indicated by narrowing of the LOA.

Discussion

The aim of the present study was to define the optimal parameters to derive sleep feature estimates from actigraphy as compared to PSG in ID and matched good sleepers. Previous studies have evaluated algorithm parameter tuning to improve the estimation of sleep features using actigraphy [11–18]. These studies evaluated a limited range of parameter settings and did not include the *sleep–wake* algorithm for sleep onset estimation, nor cross-validated the suggested optimal parameters. To overcome these limitations, we here performed a comprehensive evaluation of a wide range of parameters. We moreover also included the *sleep–wake* algorithm for sleep onset estimation, and cross-validated the optimal parameters in independent samples.

The current findings indicate that sleep–wake discrimination and sleep onset estimation are worse in ID than in CTRL, and that parameters can be chosen better than the current defaults to improve the quality of the estimates for specific populations, especially for people with ID.

The commonly used default *immobile–mobile* algorithm with a PD of 10 min achieves optimal sleep onset estimation in good sleepers and suboptimal results in ID [8–11, 13, 15]. In ID, optimal sleep onset estimation can be achieved by using the *sleep–wake* algorithm, a PD of 5 min, and a WST of 40. This setting has not commonly been reported in the literature, suggesting common use of suboptimal parameters in studies on insomnia [11, 15]. A PD of 5 min has previously been found to be optimal for sleep onset estimation with the *immobile–mobile* algorithm in patients with OSA [16].

Our findings are also in line with the notion that ID are more likely than controls to be awake while lying still in bed. Therefore, a lower, more sensitive WST may be required for ID. Indeed, a WST of 40 is optimal for TST estimation in ID, compared to the optimal WST of 90 in CTRL. The optimal WST for ID is however higher than the very low WSTs (10–20) previously suggested [11–14]. The optimal WST for CTRL is higher than the default (40) used sometimes in CTRL and ID [9], yet close to a WST of 80 used in young adults with ID [15].

The optimal parameters for the estimation of SOL and of TST are very similar in ID. In CTRL, however, the optimal estimation of SOL requires quite different parameters than the optimal estimation of TST. A single parameter configuration that achieves optimal estimates in ID and near-optimal estimates in CTRL can be used in studies where the diagnosis of ID is uncertain or changes over time, for example, as a result of intervention. To achieve optimal results within a good sleeper sample, each feature can be obtained using a distinct set of parameters. A single set of parameters will result in a trade-off of better estimates of some sleep features and worse estimates of the other sleep features.

Near-optimal estimates of sleep features could be obtained across a range of parameter configurations. Within this range, cross-validation errors remain within one standard error of the optimal configuration (Figures 1 and 2). This suggests that using parameter configurations within this range will minimally affect the accuracy of sleep feature estimates, and that it

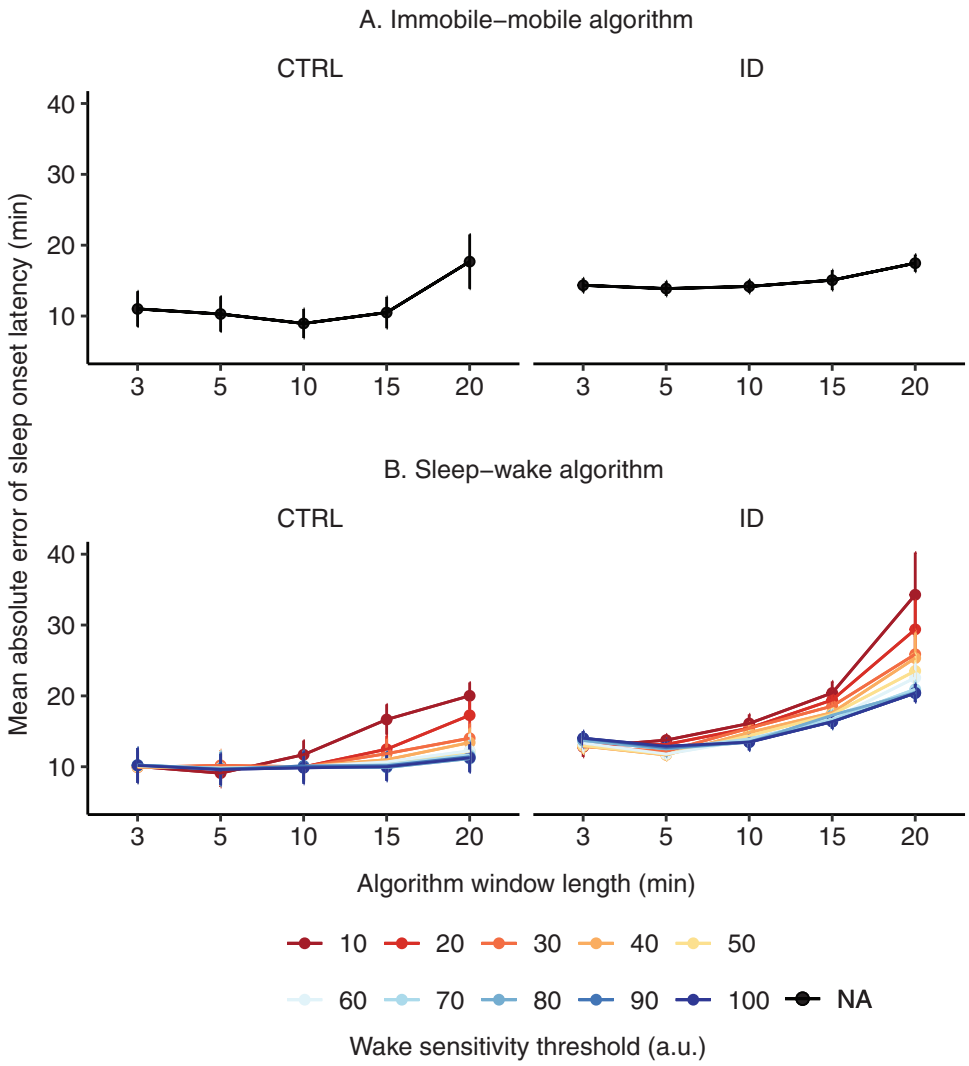

Figure 1. The MAE and standard error of the differences between polysomnographic SOL and actigraphic estimate of SOL across parameter configurations. The MAE is plotted against the PD with separate colored lines indicating different WSTs used to estimate sleep onset with (A) the *immobile–mobile* algorithm and (B) the *sleep–wake* algorithm. Since the WST does not influence the *immobile–mobile* algorithm, identical results are obtained for all WSTs and plotted as a single line in (A). Results are plotted separately for ID (right) and matched good sleepers (CTRL, left).

is reasonably safe to pool data across studies that have applied different parameter configurations.

Our findings can be generalized to actigraphs that use the algorithm implemented in the Actiware software (like Actiwatch Spectrum and Light) or to actigraphs similar to the Actiwatch used in the validation of the conversion of accelerometry to activity counts. The findings likely also generalize to other movement recordings using MEMS accelerometers similar to GENEActiv for which the same conversion steps can be applied to obtain activity counts (e.g. linear accelerometry in the palmar-dorsal axis, with a sampling frequency >30 Hz and a measurement range >5 g).

Despite the fact that sleep estimates can benefit from the optimization of parameter configuration, the agreement between PSG and actigraphy is still only moderate at best, and LOAs are relatively wide. A recent meta-analysis of the literature on the validity of actigraphy in ID has provided confidence limits for acceptable differences between actigraphy and PSG when used in clinical care decisions [7]. The limits were based on the confidence intervals (CIs) of the differences between actigraphy and PSG across multiple studies, rather than the absolute differences. Consistent differences, with small CIs, would suggest reliable actigraphic estimated sleep features compared to PSG. Differences between actigraphy and PSG with 95% CIs of 40 min for TST, 30 min for SOL, 30 min for WASO, and 5% for SE were *a priori* considered acceptable for use in clinical care decisions. A review of the studies revealed that for TST (35.12 min) and SOL (6.78 min), but not WASO (33.22 min) and SE (7.8%), the 95% CIs were small enough to prove reliable estimates compared to PSG [7]. These CIs, however, ignore the observation in our study that the bias depends on the measurement range and may therefore not be applicable to the entire range of measurements. Future recommendations may want to take this variability into account.

Innovative technology assessing other behavioral or physiological signals may be required to achieve better sleep–wake discrimination, let alone differentiation between sleep stages. Substantial discrepancies between actigraphy and PSG in some

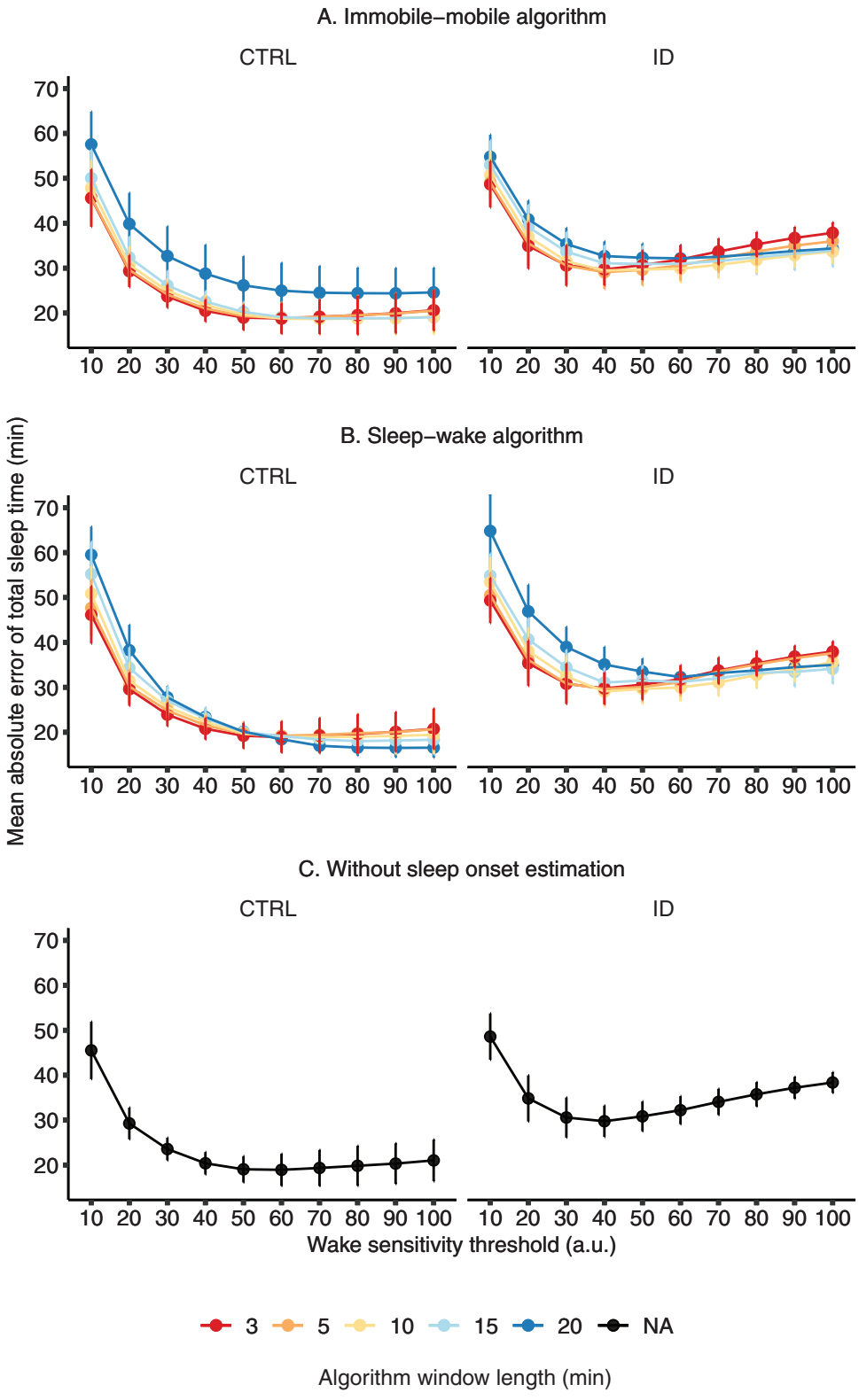

Figure 2. The MAE and standard error of the differences between polysomnographic TST and actigraphic estimate of TST across parameter configurations. The MAE is plotted against the WST with separate colored lines indicating different PDs used to estimate sleep onset with (A) the *immobile–mobile* algorithm and (B) the *sleep–wake* algorithm, or (C) no algorithm to estimate sleep onset. Since PD only influences sleep onset estimation, identical results are obtained for all PDs when sleep onset estimation is ignored and plotted as a single line in (C). Results are plotted separately for ID (right) and matched good sleepers (CTRL, left).

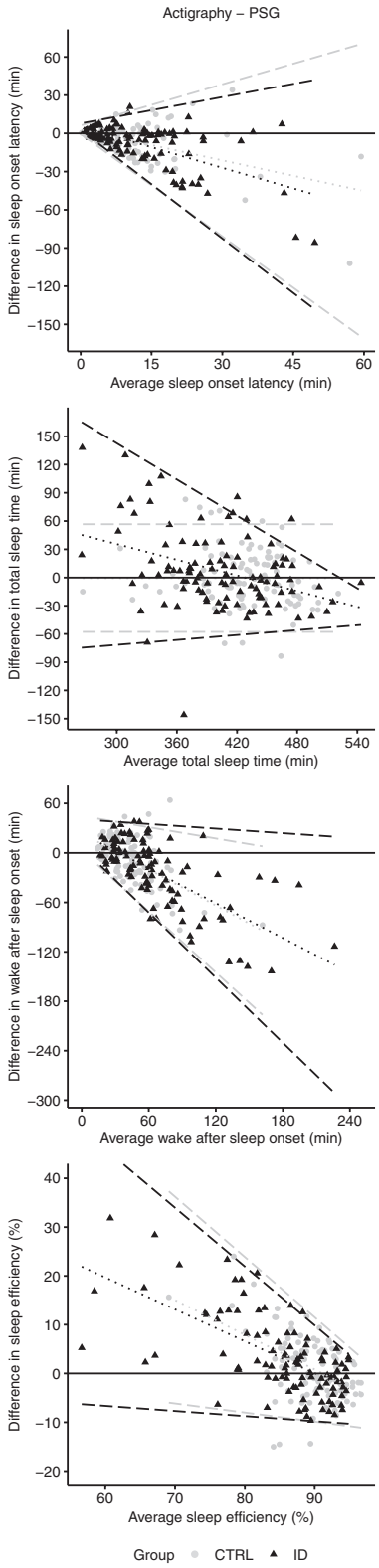

Figure 3. Bland–Altman (mean-difference) plots for SOL (top), TST, WASO, and SE (bottom) estimated with the across-group optimal *sleep–wake* algorithm, a PD of 5 min, and a WST of 40. Differences between PSG and actigraphy are plotted against the average of the two measurements for ID (triangles) and matched good sleepers (CTRL; circles). The bias (dotted lines) is the mean difference between both measurements. The 95% LOAs (dashed lines) define the range within which 95% of the differences between measurements of PSG and actigraphy will lie. The smaller the LOA, the better the agreement between both measurements. The bias and LOAs are plotted separately for ID (black) and CTRL (gray). Note the difference in the scaling of the axes.

individuals at higher WASO [35] and lower TST suggest that actigraphy may benefit from individually tuned parameters, but implementation of such may be impractical because it requires PSG.

Some limitations deserve mention. First, our study did not include sleep diaries and it should be noted that the agreement between sleep features may be higher than in real-life applications, due to the fact that self-chosen individual habitual lights-off and lights-on times were tightly aligned with the investigator-initiated start and end of the PSG recording. In clinical and ambulatory conditions, the estimation of lights-off and lights-on times can be supported by sleep diaries or light sensors. In addition to providing an in-bed interval, sleep diaries provide clinicians with valuable information regarding subjective sleep experience, making sleep diaries the current standard for assessing sleep in clinical and especially ambulatory settings. Additional studies that simultaneously assess ambulatory PSG and actigraphy would be required to evaluate the validity of diaries and light sensors in defining lights-off and lights-on times in the field. Furthermore, recent development has shown promising results for estimating the in-bed interval using a data-driven approach [36]. It would be extremely valuable if these approaches could be extended to clinical populations such as people with ID.

Second, our study design did not allow us to evaluate the performance of the algorithms to estimate sleep end, because the end of the PSG recording was tightly aligned with the wake time of the participant, with snooze duration essentially equal to 0. Sleep onset and sleep end estimation use nearly identical algorithms (with the exception that for immobility, two epochs instead of one with *any* movement are commonly allowed). Studies that simultaneously assess PSG and actigraphy while allowing unrestricted wake-up time would be required to evaluate how the currently defined optimal parameter settings perform in estimating sleep end.

Third, PSG suffers from suboptimal inter-rater agreement in sleep stage scoring. For example, microstructures from two or more sleep stages or wakefulness can occur within a 30-s epoch, and raters may differently classify the epoch depending on their judgment of the relative contribution of the microstructures within the epoch. This lack of agreement carries over to the comparison with actigraphy, as no higher agreement of actigraphy and PSG can be obtained on wakefulness and sleep than the agreement of two raters on wakefulness and sleep in PSG. Current and future development in continuous and data-driven scoring of PSG may minimize this problem. In continuous scoring, no limitations are placed on the duration of an epoch. Instead, a stage ends when the EEG microstructure indicates a transition to a new stage. Data-driven scoring of PSG may lead to more sleep stages than defined today by the AASM [37]. Both may improve inter-rater reliability in PSG and potentially its concordance with actigraphy.

Fourth, although PSG is the gold standard for sleep evaluation, it should be noted that PSG has its own limitations which may be of particular relevance to ID. The EEG signal used by PSG mostly reflects highly synchronized cortical activity. Visual or power spectral analysis of EEG reveals only cortical and subcortical activity involving sufficient neurons with aligned dipoles. It may not be surprising that the limited representation of neuronal activity provided by EEG results in a notorious lack of agreement with subjectively experienced sleep state and sleep quality, which in people with insomnia is commonly misinterpreted as sleep state "misperception" [38].

Furthermore, our sample included people with ID that were off medication. In clinical practice, actigraphy may be recommended for treatment-resistant patients that are on medication. Medications that interfere with or restrict mobility are likely to lead to overestimation of the amount of sleep by actigraphy due to immobile wakefulness classified as sleep. Unfortunately, to the best of our knowledge, no studies have evaluated the direct effects of sleep medications on the validity of actigraphy. This would be an important goal for future studies.

In conclusion, the current study defined optimal parameter settings for actigraphic estimation of polysomnographic sleep features in ID and good sleepers. We did so by evaluating a wide range of parameters, by systematically locating in-sample optimal settings, and by validating these optimal settings in independent holdout samples. The results suggest that actigraphic estimates can be improved by the use of optimized parameter configurations. Optimized settings will aid the precision of sleep estimates in increasingly available large actigraphy datasets obtained from ubiquitous wearable devices [39].

Funding

This work was supported by Project NeuroSIPE 10738, of the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs, Agriculture and Innovation; and by the European Research Council (ERC) grant ERC-2014-AdG-671084 INSOMNIA.

Disclosure Statement

Financial discloure: None declared.
Non-financial disclosure: None declared.

References

1. Wittchen HU, *et al*. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol*. 2011;**21**(9):655–679.
2. American Psychiatric Association. *DSM-5: Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Press; 2013.
3. Diagnostic Classification Steering Committee. *ICSD3—International Classification of Sleep Disorders: Diagnostic and Coding Manual*. 3rd ed. Rochester, MN: American Sleep Disorders Association; 2014.
4. Smith MT, *et al*. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American academy of sleep medicine clinical practice guideline. *J Clin Sleep Med*. 2018;**14**(7):1231–1237.
5. van Someren EJW, *et al*. A new actigraph for long-term registration of the duration and intensity of tremor and movement. *IEEE Trans Biomed Eng*. 1998;**45**(3):386–395.
6. Van Someren EJW. Improving actigraphic sleep estimates in insomnia and dementia: how many nights? *J Sleep Res*. 2007;**16**(3):269–275.
7. Smith MT, *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.
8. Sivertsen B, *et al*. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep*. 2006;**29**(10):1353–1358.
9. Sánchez-Ortuño MM, *et al*. 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.
10. McCall C, *et al*. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *J Sleep Res*. 2012;**21**(1):122–127.
11. Taibi DM, *et al*. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. *J Clin Sleep Med*. 2013;**9**(3):217–225.
12. Lichstein KL, *et al*. Actigraphy validation with insomnia. *Sleep*. 2006;**29**(2):232–239.
13. Kapella MC, *et al*. Actigraphy scoring for sleep outcome measures in chronic obstructive pulmonary disease. *Sleep Med*. 2017;**37**:124–129.
14. Natale V, *et al*. The role of actigraphy in the assessment of primary insomnia: a retrospective study. *Sleep Med*. 2014;**15**(1):111–115.
15. Williams JM, Taylor DJ, Slavish DC, *et al*. Validity of actigraphy in young adults with insomnia. *Behav Sleep Med*. 2020;**18**(1):91–106.
16. Chae KY, *et al*. Evaluation of immobility time for sleep latency in actigraphy. *Sleep Med*. 2009;**10**(6):621–625.
17. Meltzer LJ, *et al*. Comparison of actigraphy immobility rules with polysomnographic sleep onset latency in children and adolescents. *Sleep Breath*. 2015;**19**(4):1415–1423.
18. Boyne K, *et al*. Accuracy of computer algorithms and the human eye in scoring actigraphy. *Sleep Breath*. 2013;**17**(1):411–417.
19. Wei Y, *et al*. Sleep stage transition dynamics reveal specific stage 2 vulnerability in insomnia. *Sleep*. 2017;**40**(9).
20. Wassing R, *et al*. Haunted by the past: old emotions remain salient in insomnia disorder. *Brain*. 2019;**142**(6):1783–1796.
21. Benjamins J, Migliorati F, Dekker K, *et al*. Insomnia heterogeneity: Characteristics to consider for data-driven multivariate subtyping. *Sleep Med Rev*. 2017;**36**:71–81.
22. Bastien CH, *et al*. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med*. 2001;**2**(4):297–307.
23. Medicine AAoS. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: AASM; 2014.
24. Edinger JD, *et al*. Derivation of research diagnostic criteria for insomnia: report of an American Academy of Sleep Medicine Work Group. *Sleep*. 2004;**27**(8):1567–1596.
25. Iber C, *et al*. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
26. te Lindert BH, *et al*. Sleep estimates using microelectromechanical systems (MEMS). *Sleep*. 2013;**36**(5):781–789.
27. Oakley NR. Validation with polysomnography of the sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system. *Technical Report to Mini-Mitter Co Inc*. 1997.

28. Kushida CA, *et al*. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;**2**(5):389–396.
29. Gale J, *et al*. Statistical artifact in the validation of actigraphy. *Sleep*. 2005;**28**(8):1017–1018.
30. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;**20**(1):37–46.
31. Bland JM, *et al*. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;**1**(8476):307–310.
32. Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol*. 2010;**37**(2):143–149.
33. *R: A Language and Environment for Statistical Computing* [Computer program]. Vienna: R Foundation for Statistical Computing; 2018.
34. Landis JR, *et al*. The measurement of observer agreement for categorical data. *Biometrics*. 1977;**33**(1):159–174.
35. Marino M, *et al*. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*. 2013;**36**(11):1747–1755.
36. van Hees VT, *et al*. Estimating sleep parameters using an accelerometer without sleep diary. *Sci Rep*. 2018;**8**(1):12975.
37. Christensen JAE, *et al*. Data-driven analysis of EEG reveals concomitant superficial sleep during deep sleep in insomnia disorder. *Front Neurosci*. 2019;**13**:598.
38. te Lindert BHW, *et al*. Actigraphic multi-night home-recorded sleep estimates reveal three types of sleep misperception in Insomnia Disorder and good sleepers. *J Sleep Res*. 2020;**29**(1):e12937.
39. Khosla S, *et al*. Consumer sleep technology: an American academy of sleep medicine position statement. *J Clin Sleep Med*. 2018;**14**(5):877–880.