RESEARCH ARTICLE

Validation of the Munich Actimetry Sleep Detection Algorithm for estimating sleep–wake patterns from activity recordings

Ann-Sophie Loock[1] | Ameena Khan Sullivan[1,2] | Catia Reis[1,3,4] 🆔 | Teresa Paiva[3,4] | Neda Ghotbi[1,5] 🆔 | Luisa K. Pilz[1,6] 🆔 | Anna M. Biller[1,7] | Carmen Molenda[1] | Maria T. Vuori-Brodowski[1] | Till Roenneberg[1] 🆔 | Eva C. Winnebeck[1] 🆔

[1]Institute of Medical Psychology, Ludwig Maximilian University Munich, Munich, Germany

[2]School of Biosciences and Medicine, University of Surrey, Guildford, UK

[3]ISAMB – Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

[4]CENC – Sleep Medicine Center, Lisboa, Portugal

[5]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital, Ludwig Maximilian University Munich, Munich, Germany

[6]Laboratório de Cronobiologia e Sono (HCPA/UFRGS)/Graduate Program in Psychiatry and Behavioral Sciences (Universidade Federal do Rio Grande do Sul), Porto Alegre, Brazil

[7]Graduate School of Systemic Neurosciences, Ludwig Maximilian University Munich, Munich, Germany

Correspondence
Till Roenneberg and Eva C. Winnebeck, Institute of Medical Psychology, Goethestraße 31, 80336 Munich, Germany.
Emails: till.roenneberg@med.uni-muenchen. de (TR), eva.winnebeck@med.uni-muenchen. de (EW)

Present address
Till Roenneberg, Institute and Polyclinic for Occupational, Social and Environmental Medicine, Ludwig Maximilian University Munich, Munich, Germany

Funding information
ASL received a stipend from the Max-Weber-Programm (Studienstiftung), AMB received funding from the Graduate School of Systemic Neurosciences Munich, CR received funding from the Fundação para a Ciência e Tecnologia (FCT) PhD research grants (PDE/BDE/114584/2016), LKP received a fellowship from the Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES, Finance Code 001), and NG received research funding from the FoeFoLe program at LMU (registration No. 37/2013).

Abstract

Periods of sleep and wakefulness can be estimated from wrist-locomotor activity recordings via algorithms that identify periods of relative activity and inactivity. Here, we evaluated the performance of our Munich Actimetry Sleep Detection Algorithm. The Munich Actimetry Sleep Detection Algorithm uses a moving 24–h threshold and correlation procedure estimating relatively consolidated periods of sleep and wake. The Munich Actimetry Sleep Detection Algorithm was validated against sleep logs and polysomnography. Sleep-log validation was performed on two field samples collected over 54 and 34 days (median) in 34 adolescents and 28 young adults. Polysomnographic validation was performed on a clinical sample of 23 individuals undergoing one night of polysomnography. Epoch-by-epoch analyses were conducted and comparisons of sleep measures carried out via Bland-Altman plots and correlations. Compared with sleep logs, the Munich Actimetry Sleep Detection Algorithm classified sleep with a median sensitivity of 80% (interquartile range [IQR] = 75%–86%) and specificity of 91% (87%–92%). Mean onset and offset times were highly correlated ($r = .86$–.91). Compared with polysomnography, the Munich Actimetry Sleep Detection Algorithm reached a median sensitivity of 92% (85%–100%) but low specificity of 33% (10%–98%), owing to the low frequency of wake episodes in the night-time polysomnographic recordings. The Munich Actimetry Sleep Detection

2 of 12

Journal of
Sleep
Research

ESRS

LOOCK ET AL.

Algorithm overestimated sleep onset (~21 min) and underestimated wake after sleep onset (~26 min), while not performing systematically differently from polysomnography in other sleep parameters. These results demonstrate the validity of the Munich Actimetry Sleep Detection Algorithm in faithfully estimating sleep–wake patterns in field studies. With its good performance across daytime and night-time, it enables analyses of sleep–wake patterns in long recordings performed to assess circadian and sleep regularity and is therefore an excellent objective alternative to sleep logs in field settings.

1 | INTRODUCTION

Sleep is characterized by reduced or absent consciousness, perceptual disengagement, immobility and a characteristic sleep posture (Grandner & Rosenberger, 2019). The current gold standard to objectively detect and monitor sleep in humans is polysomnography (PSG), based on characteristic electroencephalogram (EEG) patterns, muscle tone and eye movements. Despite its usefulness in many clinical and research settings and its rich detail, PSG recordings and analyses are laborious and expensive, and thus rarely performed over several nights on the same person, and consequently not well suited to studying habitual sleep–wake patterns (Grandner & Rosenberger, 2019; Van de Water et al., 2011).

In turn, actigraphy or actimetry (as we prefer to call it) monitors states of immobility through the detection of movements by wrist-worn devices (Conley et al., 2019; Dick et al., 2010; Grandner & Rosenberger, 2019; Marino et al., 2013; Roenneberg et al., 2015; Toon et al., 2016). Pioneering work in the 1970s and 1980s has made long, continuous recordings possible and demonstrated that sleep times and duration can be estimated from such records by identifying periods of relative immobility (Borbély, 1986; Kripke et al., 1978). Thus, sleep–wake patterns can be captured over days, months or even years (Borbély et al., 2017) and analysed for sleep regularity, weekly or even seasonal patterns, effects of interventions, or sleep–wake rhythm disturbances resulting from circadian rhythm disorders (Ancoli-Israel et al., 2003; Dick et al., 2010; Kantermann et al., 2007; Roenneberg et al., 2015; Sadeh, 2011; Smith et al., 2018). In addition to providing long-term records of sleep–wake patterns that are a prerequisite for circadian analyses, actimetry is less expensive and less sleep disturbing than PSG and captures sleep where it normally occurs (Ancoli-Israel et al., 2003; Conley et al., 2019; Dick et al., 2010; Marino et al., 2013; Toon et al., 2016; Tryon, 2004). The easy handling and at-home applicability may also result in higher participation rates than in PSG studies (Marino et al., 2013). Actimetry even poses an advantage over sleep logs, which can be cumbersome to fill in over a long period and require literacy, but have often been the only practical alternative to investigate the long-term structure of sleep–wake rhythms (Girschik et al., 2012).

The performance of actimetry in comparison to sleep logs and PSG in sleep–wake detection depends on the sleep–wake scoring algorithm, the recording device, the study population and, of course, the question at hand. Generally, validations against PSG indicate adequate estimation of time and duration of sleep episodes as long as individuals do not have severe sleep fragmentation or severe sleep disorders (Ancoli-Israel et al., 2003; Sadeh, 2011). In such studies, actimetry was repeatedly shown to be highly sensitive in detecting sleep (sensitivity) but quite insensitive in detecting wake (specificity), thus indicating a tendency for movement-based sleep–wake scoring to overestimate sleep and underestimate wake (Conley et al., 2019; Grandner & Rosenberger, 2019; Marino et al., 2013; Sadeh, 2011; a et al., 2011). Importantly, these studies were mainly based on night-time recordings, where sleep (not wake) is the most abundant and probable state. Therefore, the underestimation of wake reflects the underestimation of wake close to sleep onset and of wake disruptions of sleep (i.e., wake after sleep onset [WASO]), when people may are relatively immobile in their beds- and not necessarily the underestimation of wakefulness during the day when people tend to move more. Accordingly, in validations against sleep logs that included wakefulness during the day, actimetric sleep–wake detection performed well in both sleep and wake detection and thus allowed monitoring changes in sleep patterns over time (Iwasaki et al., 2010; Lockley et al., 1999; Santisteban et al., 2018).

Here, we present the validation of our sleep–wake scoring algorithm called Munich Actimetry Sleep Detection Algorithm (MASDA; sometimes previously referred to as bin-sleep method; Roenneberg et al., 2015). Like the many other algorithms introduced since the first validated algorithm by Webster et al., (1982), MASDA weighs the movement values within an epoch of interest against previous and subsequent epochs (Grandner & Rosenberger, 2019). However, the MASDA was specifically designed from a circadian perspective to prioritize the detection of general sleep–wake patterns over the detection of sleep–wake in each individual short epoch. Hence it operates on a 10-min resolution of movement counts (10-min analysis epochs), which we commonly use in circadian actimetry analyses (Roenneberg et al., 2015), employs a 24-h moving threshold for primary

sleep–wake detection and yields relatively consolidated sleep episodes via a secondary, dedicated correlation procedure. This was intended to enable easier analyses and better pattern recognition on a 24-h scale and long-range recordings than the more common short epoch-by-epoch approach used in most standard algorithms (e.g. Cole-Kripke's, Scripps', Oakley's, Sadeh's) (Cole et al., 1992; Kripke et al., 2010; Oakley, 1997; Sadeh et al., 1994).

In this study, we validated the sleep–wake scoring results of the MASDA against sleep-log entries from two samples collected over multiple weeks in the field as well as against PSG during single laboratory nights in a clinical sample.

2 | METHODS

2.1 | Sleep log sample

2.1.1 | Participants

For the validation against sleep logs, we used two samples from previous studies, an adolescent sample and one young adult sample. The adolescent sample was collected over 9 weeks in 45 German high-school students (mainly Caucasians), of whom 34 participants (22 females, mean [M] = 16.7 years, standard deviation [SD] = 1.2 , range = 14–19) provided high-quality data in both their sleep logs and actimetry records (median of 54 days) and were used for further analyses (Winnebeck et al., 2020). The young adult sample was collected over 4–6 weeks in 30 German participants (mainly Caucasians), of whom 28 (13 females, M = 22.8 years, SD = 3.6, range = 19–33) provided complete data across both methods (median of 34 days) and made up the final adult sample (Ghotbi et al., 2020). Approval for both studies was obtained by the Ethics Committee of the LMU Medical Faculty (517–15, 774–16), and all participants (and their guardians if applicable) provided informed consent.

2.1.2 | Actimetry

Activity was recorded with wrist-worn devices (Daqtometer 2.4, Daqtix) that were worn continuously on the wrist of the dominant or non-dominant hand (participants' choice). This choice was possible as we did not aim to estimate general physical activity for metabolic monitoring but to estimate activity patterns. These dual-axis accelerometers were set to sample static and dynamic acceleration at 1 Hz. Activity counts (the sum of the linear differences of subsequent readings for each axis) were stored by the devices at 30-s intervals as the mean of all counts in this interval.

2.1.3 | Sleep log

The sleep logs for both samples were based on the μMCTQ (Ghotbi et al., 2020), a short version of the Munich ChronoType Questionnaire. Instead of asking participants to record their average sleep times for the last weeks separately for work and work-free days, as this questionnaire normally does, the μMCTQ was applied daily via an online platform (limesurvey.org) to record initial sleep onset and final sleep offset of the previous night.

2.2 | Polysomnography sample

2.2.1 | Participants

For the validation against laboratory PSG, a dedicated dataset was recorded at the CENC Sleep Medicine Center, Lisbon, Portugal. The original sample consisted of 50 participants. However, because of software and signal synchronization issues (n = 11 and 16, respectively), records from only 23 of these participants (9 females, M = 40.1 years, SD = 13.7, range = 21–80) could be used for analysis. Of these, 11 subjects were diagnosed based on the PSG recording as without any clinical sleep pathology (healthy), four with insomnia, one with parasomnia (rapid eye movement [REM] Behaviour Disorder), four with circadian rhythm sleep–wake disorder (two delayed sleep–wake phase disorder and two shift work disorder), and three with sleep related breathing disorder (obstructive sleep apnea). The study was approved by the Lisbon Medical School Ethics Committee and all participants gave their written consent.

2.2.2 | Actimetry

Participants wore wrist actimeters ≥24 h before and after the laboratory PSG night for a total of 14 days. For the majority of participants, ActTrust devices (Condor Instruments) were used; for two participants the Actiwatch 2 (Phillips Respironics) was used. Sensitivity analyses without data from Actiwatch-2 participants yielded equivalent results, indicating that device differences did not drive results. For both devices, activity was sampled every second and stored in 30-s bins.

Adequate temporal synchronization between PSG and actimetry was established via event markers in both recordings: a marker button signal for the actimeters and the lights-off signal for the PSG. When there was a mismatch of >10 min between the marker-time stamps (indicative of an error at the time of recording rather than a temporal mismatch between the devices), manual matching via the actimetric light profile was attempted, and otherwise the records excluded from analysis (n = 16).

2.2.3 | Polysomnography

Overnight PSG was performed with the Nicolet System (Viasys Healthcare) or the Domino Somnoscreen Plus (Somnomedics). The recorded parameters included: electroencephalography (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1); left and right

electrooculogram; submental electromyogram; bilateral tibial electromyogram; electrocardiogram; oronasal airflow with three-pronged thermistors; nasal pressure with a pressure transducer; rib cage and abdominal wall motion via respiratory impedance plethysmography; arterial oxygen saturation with pulse waveform; and digital video and audio. The recording was scored from "lights off" to "lights on," with lights off scheduled as close as possible to participants' normal sleep schedule, aiming for a sleep period of 8 h. In the case of an interfering work schedule or a sleep disorder that prevented participants from staying asleep (e.g., insomnia), a sleep period of fewer than 8 h was tolerated. The recordings were manually scored in 30-s epochs by trained sleep technicians according to the American Academy of Sleep Medicine specifications (AASM version 2.3, 2016).

To match the 10-min resolution underlying the MASDA (see below), the PSG scoring at 30-s resolution was aggregated into 10-min intervals via the mode (i.e., the most prevalent sleep stage over 20 consecutive 30-s epochs was assigned to each 10-min interval). These were subsequently converted to a binary categorization (0 = wake, 1 = sleep). The median length of the series was 47.0 10-min intervals (IQR = 42.5–51.0) (i.e., 7.8 h). Participants spent 41.0 (37.5–44.5) intervals asleep and 5.0 (2–8.5) intervals awake.

2.3 | The Munich Actimetry Sleep Detection Algorithm

The Munich Actimetry Sleep Detection Algorithm (Roenneberg et al., 2015), formerly also referred to as bin sleep method, is a two-step procedure for binary sleep–wake scoring from activity counts, heuristically designed to yield relatively consolidated stretches of sleep or wake. If it is desired to use it with raw accelerometry data, the data have to be converted to counts prior to analyses, using approaches such as the te Lindert method (te Lindert & Van Someren, 2013).

The first step of sleep detection is a threshold procedure in which all epochs (usually 10 min long) with activity counts below a given percentage of the 24-h-centred moving average are classified as putative sleep. The default percentage we use is 15%, but it can be adapted from 10%–25% for specific populations or individuals with particularly low or high activity during wake or sleep. The second step of the MASDA is a "cleaning" procedure consisting of a duration filter and a correlation procedure. The filter reclassifies any sleep epoch not part of a stretch of at least 30 min as wake to avoid misclassification of short periods of inactivity. This is followed by a correlation procedure that joins adjacent stretches of sleep epochs based on a test series of sleep episodes of varying lengths. For more details on the correlation procedure, please refer to Roenneberg et al. (2015). The algorithm was originally implemented in C++ in the software ChronoSapiens (© 2020 Chronsulting UG; Roenneberg et al., 2015), but has also been included in the Python package pyActigraphy (Hammad et al., 2020).

2.4 | Activity data processing

Actimetry data were analysed via our in-house software ChronoSapiens (Version 10). All activity records were analysed via our standard 10-min resolution (Roenneberg et al., 2015); the data were aggregated into intervals of 10 min via the arithmetic mean upon import into the program. Periods of non-wear were identified based on participants' self-reports (actimetry logs) as well as based on stretches of consecutive zeros exceeding 100 min and excluded from the analysis (i.e., set to "not available" [NA]). If these stretches occurred at the beginning of the inactive period on multiple days in the same individual, these were taken to be sleep with hardly any movement and not replaced with NA.

Sleep detection via MASDA was performed with a 15% threshold (20% for the adolescent sample), and the setting to perform correlation series for four 10-min bins past the last r_{max} was 4. The 15% and 4-r_{max}-settings are our default settings and should be the starting point for any algorithm tuning to a specific population. We suggest tuning the threshold between 10% and 25% and the r_{max} between 3 and 5. The difference between a 15% and a 20% threshold is negligible in most populations and usually only mildly affects sleep–wake detection in a few individuals. We therefore advise starting with the default settings and, in the case of an obvious mismatch between visual and algorithm-based sleep detection in some individuals, to change first the threshold across all participants to identify if there is a better setting for the entire sample (reason for the 20% setting in our adolescent sample). If specific individuals cannot be accommodated with a general population setting, one can also consider participant-specific settings, particularly in diverse or severely sleep-disrupted samples (e.g., shift workers; Vetter et al., 2015).

Because the MASDA incorporates information from the surrounding 24 h via the 24-h moving average, it can be influenced, to a certain extent, by stretches of missing data. To avoid systematic effects on the sleep–wake scoring, sleep bouts in the sleep-log samples, where we can afford stricter criteria given the large amount of data, were excluded from the analysis (i.e., epochs set to NA) if (a) ≥1 h of missing data was present within 3 h or (b) ≥4 h of missing data was present within 15 h before or after the sleep bout. Any sleep bouts within the first or last 15 h of the recording were also excluded. This led to a mean proportion of 3.5% (SD 3.0) epochs of NA per sleep-log record.

2.5 | Method comparisons

2.5.1 | Epoch-by-epoch agreement

Sensitivity, specificity, predictive values and overall accuracy served as measures of agreement between the MASDA and sleep logs/PSG. Sensitivity means the proportion of "true" sleep epochs (according to sleep log/PSG) that are also identified as sleep by the MASDA. Specificity is defined as the proportion of "true" wake epochs (according to sleep log/PSG) that are also rated as wake by the MASDA.

LOOCK ET AL.

Journal of
Sleep
Research

ESRS

5 of 12

Whereas sensitivity and specificity relate the classification of the MASDA to the ground truth (sleeplogs/PSG), the predictive values describe the probability that a classification obtained by the MASDA is correct, taking the relative prevalence of sleep versus wake into account: the positive predictive value (PPV) quantifies accurate ratings of sleep; the negative predictive value (NPV) quantifies accurate ratings of wake. Overall accuracy is defined as the proportion of all sleep log/PSG bins that are correctly classified by the MASDA. Analyses were based on pairwise-complete epochs; that is, epochs with invalid or missing data (NA) in either of the two methods under comparison were disregarded during the analysis.

2.5.2 | Sleep parameter agreement

Summary sleep parameters calculated from the epoch time courses included sleep onset, offset and duration (called total sleep time, TST, for the PSG sample) for both the sleep log and PSG samples, as well as sleep period time (SP), WASO, and sleep efficiency (SE) for the PSG sample. For the sleep-log sample, which encompassed multiple weeks of recordings, the average sleep onset and offset times and durations per person over the entire recording were used. These averages were calculated after eliminating naps and fusing adjacent sleep bouts to obtain a daily onset and offset of the main sleep episode; for duration, interim wake periods were subtracted (see Winnebeck et al. (2020) for details). For the PSG sample, sleep onset was defined as the first bin scored as sleep after PSG recording started and sleep offset as the last bin scored as sleep before the PSG scoring ended. Sleep period time was defined as the elapsed time between sleep onset and sleep offset. Wake after sleep onset was calculated via the number of wake bins within a sleep episode. Total sleep time was defined as SP minus the amount of WASO. Sleep efficiency was the proportion of TST relative to time in bed (i.e., time between lights off and lights on).

Correlation analysis of sleep parameters from PSG and the MASDA was performed either via Pearson product moment correlations or via Spearman rank order correlations if parameters were non-normally distributed according to the Shapiro-Wilk test. The alpha level was set to 0.05. Additionally, Bland-Altman plots were created to visually examine the systematics of potential deviations between the sleep parameters derived from the two methods.

All analyses were conducted using R 3.5.1 and 4.0.2 (R Core Team, 2020) with special packages including psych (Revelle, 2020), tidyverse (Wickham et al., 2019) and data.table (Dowle & Srinivasan, 2020). Plots were generated using ggplot2 (Wickham, 2016) in R and matplotlib (Hunter, 2007) in Python 2.7.16 (Python Software Foundation, 2001–2019).

3 | RESULTS

For our validation of the Munich Actimetry Sleep Detection Algorithm, we made use of three different samples. Two samples

with long continuous field recordings (medians of 54 and 34 days), one of adolescent students ($n = 34$) and one of young adults ($n = 28$), provided the basis for assessing the MASDA against sleep-log records. A clinical sample with overnight PSG ($n = 23$) including both patients with various sleep disorders as well as healthy sleepers was used to assess the algorithm against PSG. Validation included both assessment of epoch-by-epoch agreement as well as comparisons of standard summary parameters.

3.1 | Validation against sleep logs

3.1.1 | Epoch-by-epoch agreement

For each individual of the adolescent and young adult samples, the sleep/wake classification for each 10-min epoch was compared between the MASDA and sleep logs. Over all participants, the MASDA reached a median accuracy of 87% (IQR = 84%–89%), sensitivity of 80% (75%–86%), specificity of 91% (87%–92%), positive predictive value of 80% (76%–85%) and a median negative predictive value of 90% (88%–92%; Figure 1a). The Munich Actimetry Sleep Detection Algorithm thus performed adequately in recognizing sleep: 80% of sleep-log-rated sleep epochs were correctly identified by the MASDA (sensitivity), and 80% of algorithm-determined sleep epochs were also rated as sleep epochs in the logs (positive predictive value). In these long, continuous recordings, the algorithm performed even better in recognizing consolidated wake epochs: sleep-log-rated wake was identified as wake by the MASDA in 91% of epochs (specificity), and 90% of epochs classified as wake by the MASDA were sleep-log-rated wake as well (negative predictive value). Importantly, we identified no systematic differences between the two samples in any of the metrics (Table S1).

3.1.2 | Agreement in summary sleep parameters

Agreement in sleep onset, offset and duration between the MASDA and sleep logs was assessed via correlations and Bland-Altman plots in the adolescent sample (Figures 2 and 3). The summary statistics of the parameters themselves are listed in Table S2.

Correlation analyses revealed strong positive associations between both methods in all three summary parameters ($R_{onset} = .92$, $R_{offset} = .86$, $R_{duration} = .62$; all $p < .001$). These associations remained strong when differentiating between schooldays ($R_{onset} = .89$, $R_{offset} = .80$, $R_{duration} = .78$; all $p < .001$) and weekends ($R_{onset} = .91$, $R_{offset} = .86$, $p < .001$), albeit with a moderate association for duration on weekends ($R = .46$, $p < .01$).

The Bland-Altman analyses, which assess potential systematic disagreements between log and MASDA-determined parameters, show that both mean sleep onset (Figure 3a) and offset times (Figure 3b) were on average 21 min earlier from MASDA. Mean sleep durations (Figure 3c) were quite similar for both methods, with MASDA-derived durations being on average 6 min shorter than
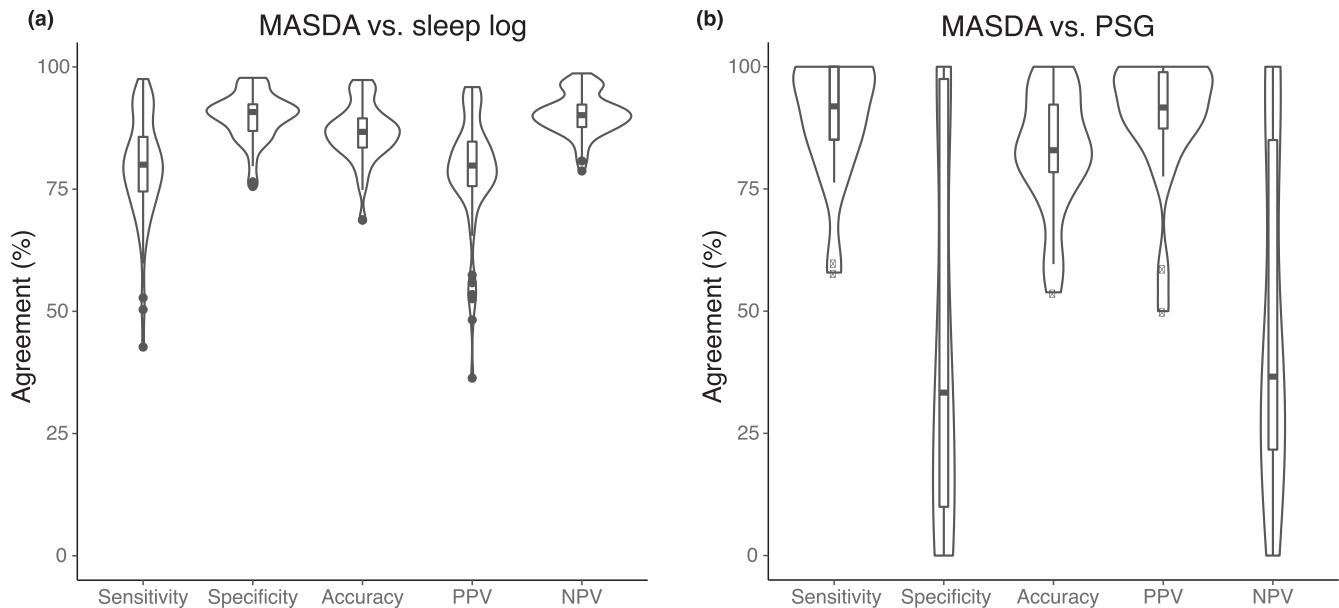
FIGURE 1 Epoch-by-epoch agreement between the Munich Actimetry Sleep Detection Algorithm (MASDA) versus sleep logs and MASDA versus polysomnography. (a) Agreement of the sleep–wake scoring between MASDA and sleep logs, with sleep log scoring as the ground truth; assessed across multiple weeks in an adolescent and a young adult sample (total $n = 62$). See Table S1 for the individual results of each sample. (b) Agreement of the sleep–wake scoring between MASDA and PSG, with PSG as the ground truth; assessed in single nocturnal recordings from a clinical sample ($n = 23$). Results are displayed as a combination of violin and Tukey boxplots to illustrate data distribution. NPV, negative predictive value; PPV, positive predictive value; PSG, polysomnography

those from logs. The differences in sleep duration between the two methods showed more variability than those of the other parameters. No systematic differences with later onset or offset times or longer durations emerged from the analyses.

3.2 | Validation against PSG

3.2.1 | Epoch-by-epoch agreement

Using the PSG sample, the sleep/wake classification of the MASDA was compared to the PSG classification for each 10-min analysis epoch within each individual. Over all participants, the MASDA reached a median accuracy of 83% (IQR = 78%–92%), sensitivity of 92% (85%–100%), specificity of 33% (10%–98%), positive predictive value of 92% (87%–99%) and negative predictive value of 37% (22%–85%; Figure 1b). Of note, specificity and negative predictive value both spanned the complete range from 0% to 100% (Figure 1b). In the PSG validation, where only night-time sleep–wake states were assessed, the MASDA performed best in detecting sleep: epochs considered sleep in the PSG scoring were identified as sleep by the MASDA in 92% of cases (sensitivity); sleep epochs identified by the MASDA were also PSG-identified sleep epochs in 92% of cases (positive predictive value). However, the MASDA showed more difficulty in detecting wake: only 33% of epochs that were considered wake by PSG were identified as wake by the MASDA (specificity), and those that were classified as wake by the MASDA were correct in only 37% of cases (negative predictive value). The lowest values for specificity

and negative predictive value were in individuals with very few PSG-determined wake epochs. Here, misclassification by the algorithm weighed particularly strongly by definition.

3.2.2 | Agreement in summary sleep parameters

In addition to the epoch-by-epoch comparisons, we also analysed agreement in common summary sleep parameters (for descriptives see Table S2). Spearman correlation analyses between the parameters of both methods revealed a strong relationship for sleep onset ($rho = 0.63$, $p = .01$; Figure 4a) and for sleep offset ($rho = 0.76$, $p < .001$; Figure 4b). In contrast, sleep period duration, TST, WASO and sleep efficiency, which more heavily depend on wake detection during the sleep episode, showed no statistically significant relationship between the actimetry-determined and the PSG-determined values (Figure 4c–g).

Furthermore, Bland-Altman plots were used for visual inspection of potential systematic disagreements between actimetry-determined and PSG-determined summary measures (Figure 5). Bland-Altman analysis of sleep onset (Figure 5a) showed that onset times from the MASDA were on average 21 min later than those from PSG. This pattern became more pronounced the later the sleep onset occurred (in relation to the start of the PSG recording). In contrast, sleep offset times for both methods (Figure 5b) were very similar, regardless of the relative timing of the sleep offset, with the bin sleep-detected offset deviating on average by less than 1 min from PSG-determined values. In accordance with these later onset
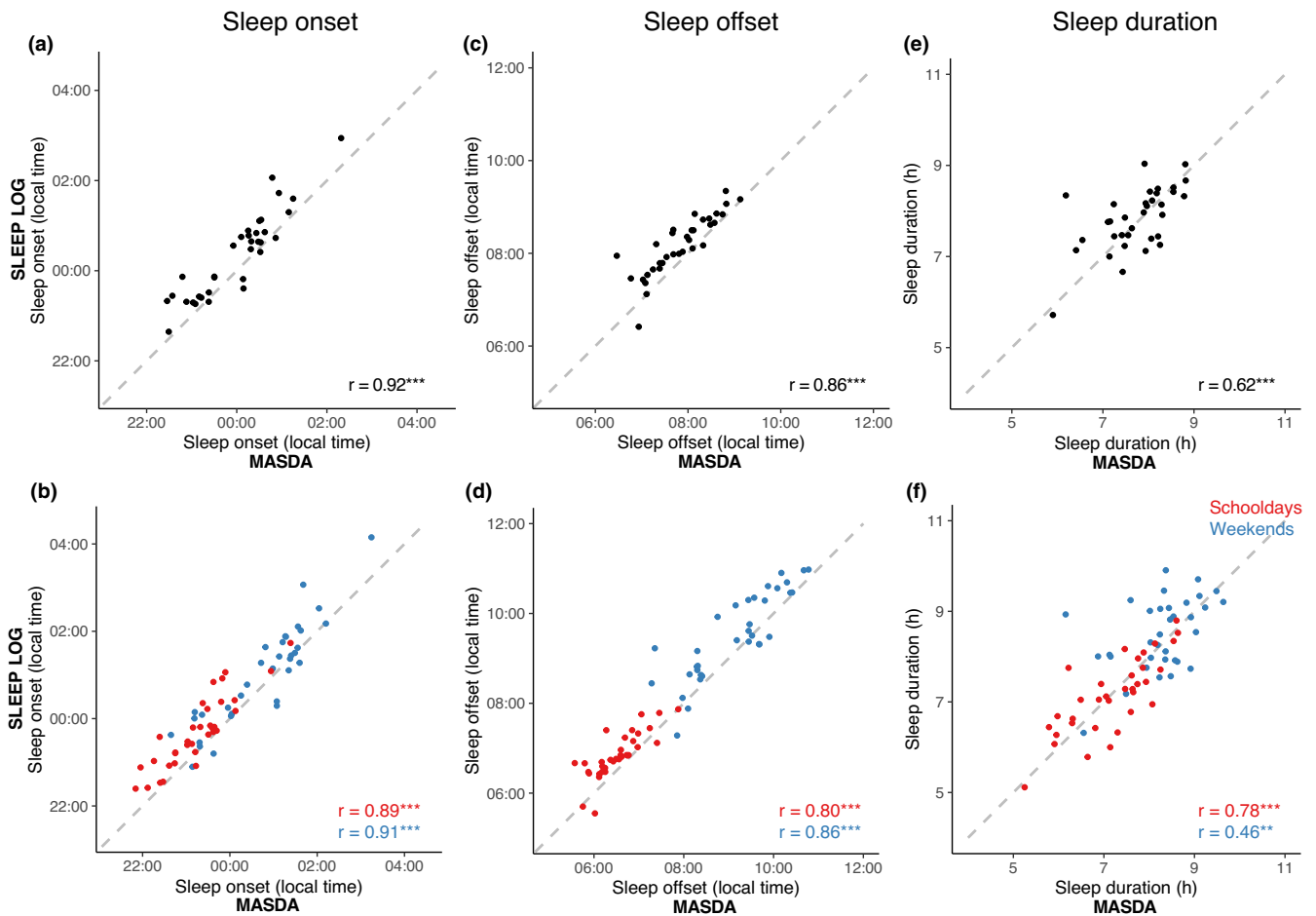
LOOCK ET AL.

Journal of
Sleep
Research

ESRS

7 of 12

FIGURE 2 Correlation of summary sleep parameters from the Munich Actimetry Sleep Detection Algorithm (MASDA) and sleep logs. Mean sleep onset times (a, b), sleep offset times (c, d) and sleep duration (e, f) of each participant from the adolescent sample (n = 34) as determined via the MASDA (x-axes) against those from sleep logs (y-axes). In panels (a), (c) and (e), means across all assessment days are compared, and in (b), (d) and (f) the comparison is differentiated into means from schooldays (red) and weekends (blue). Results of Pearson correlations are provided (**p < .01; ***p < .001); dashed line represents a 1:1 relationship

and similar offset times, the duration of the sleep period (from initial sleep onset until final sleep offset; Figure 5c) was on average 20 min shorter in the MASDA. MASDA also underestimated WASO (Figure 5d), as was expected from the lower specificity values obtained in the epoch-by-epoch analyses. Wake after sleep onset from the MASDA was on average 26 min shorter than from PSG, notably showing outliers in both directions and no obvious relationships between WASO amount and method deviance. Lastly, both TST (Figure 5e) and SE (Figure 5f) deviated on average only marginally between the two methods. MASDA equally under- and overestimated TST and SE, whereas the deviance did not show any dependency on the amount of TST and SE.

Taken together, despite differences between the two methods, the Bland-Altman plots suggest that the MASDA did not perform systematically worse than PSG in this one-night recording in most sleep parameters, except for sleep onset and WASO estimation. A few outliers in the difference between the MASDA and PSG scores can be detected in most parameters, especially from one individual with delayed sleep–wake phase disorder (DSWPD). Nonetheless,

sleep disorders do not seem to have systematically affected the MASDA results as far as can be judged from this small sample (see colour coding in Figure 5).

4 | DISCUSSION

In our comparison of the MASDA to sleep logs and PSG in three samples, we observed adequate rates of agreement throughout. In reference to sleep logs, the MASDA performed well in detecting both consolidated sleep as well as consolidated wake states in epoch-by-epoch analyses. The summary parameters sleep onset, offset and duration from the MASDA and sleep logs correlated highly, with onsets and offsets from the MASDA systematically earlier than log-derived values. In reference to a single night of PSG recordings, the MASDA correctly identified sleep in most cases, yet showed a lower performance in detecting wake (i.e., the short, more unstable states of wake right before and after sleep onset). The Munich Actimetry Sleep Detection Algorithm deviated most from PSG in the
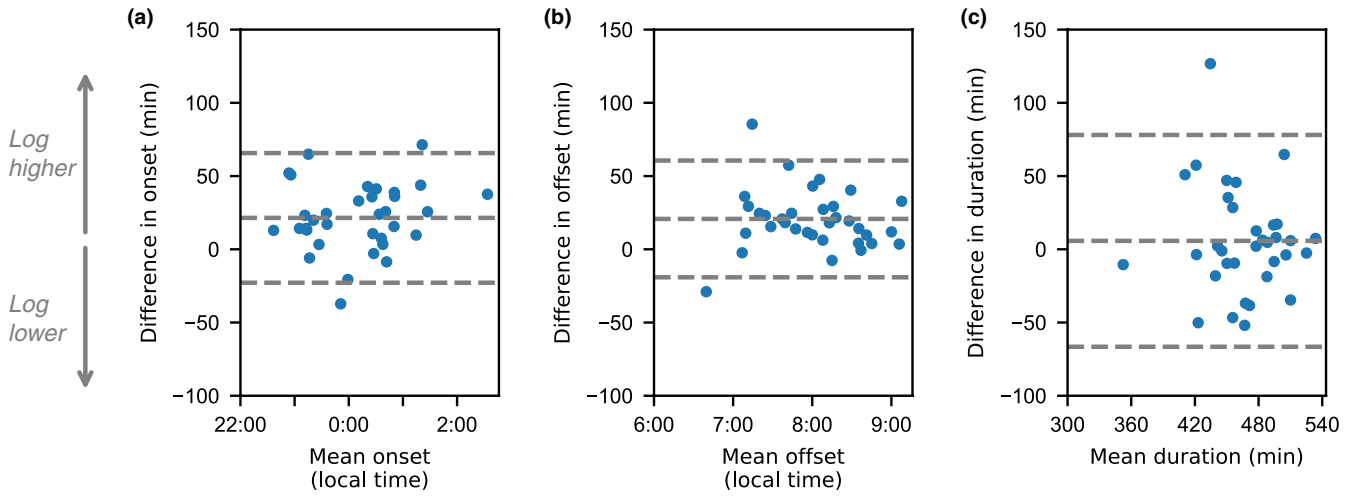
FIGURE 3 Bland-Altman analysis of summary sleep parameters from the Munich Actimetry Sleep Detection Algorithm (MASDA) and sleep logs. Bland-Altman plots for (a) sleep onset, (b) sleep offset and (c) sleep duration for the adolescent sample (*n* = 34). In each panel, the mean between the sleep log and the MASDA is plotted on the *x*-axis, the absolute difference (log-MASDA) on the *y*-axis. The mean difference is denoted by the dashed line in the middle, upper and lower boundaries of the 95% confidence interval by the upper and lower dashed lines
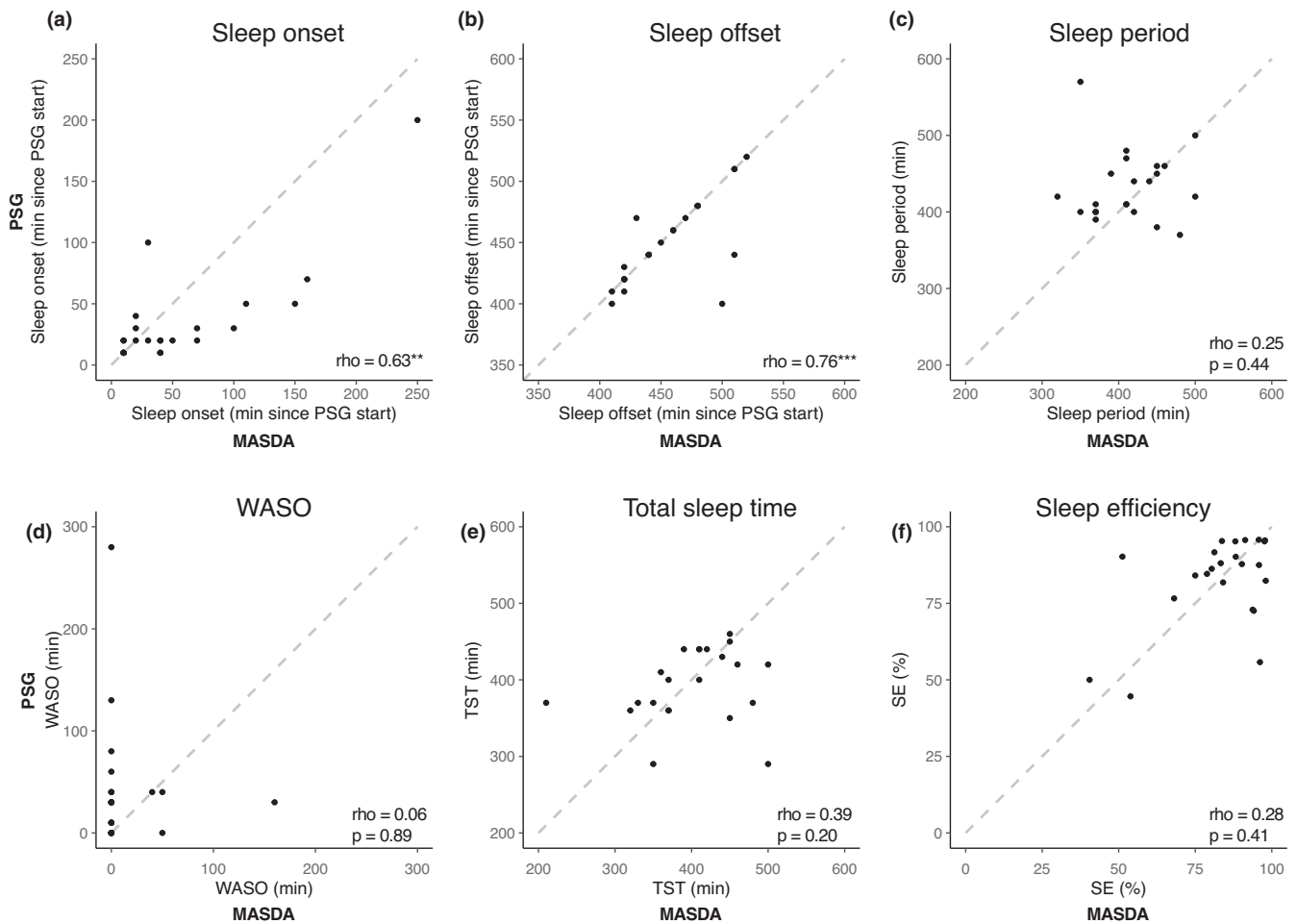

FIGURE 4 Correlation of summary sleep parameters from the Munich Actimetry Sleep Detection Algorithm (MASDA) and polysomnography. For each participant from the polysomnography (PSG) sample (*n* = 23), the MASDA-determined sleep parameter (*x*-axes) is plotted against the polysomnography-determined parameter (*y*-axes). (a) Sleep onset, (b) sleep offset, (c) sleep period, (d) WASO, (e) total sleep time, (f) sleep efficiency. Results of Spearman correlations are provided (*p*-values adjusted for multiple testing using the Benjamini-Hochberg correction; ***$p < .001$; **$p < .01$). SE, sleep efficiency; TST, total sleep time; WASO, wake after sleep onset
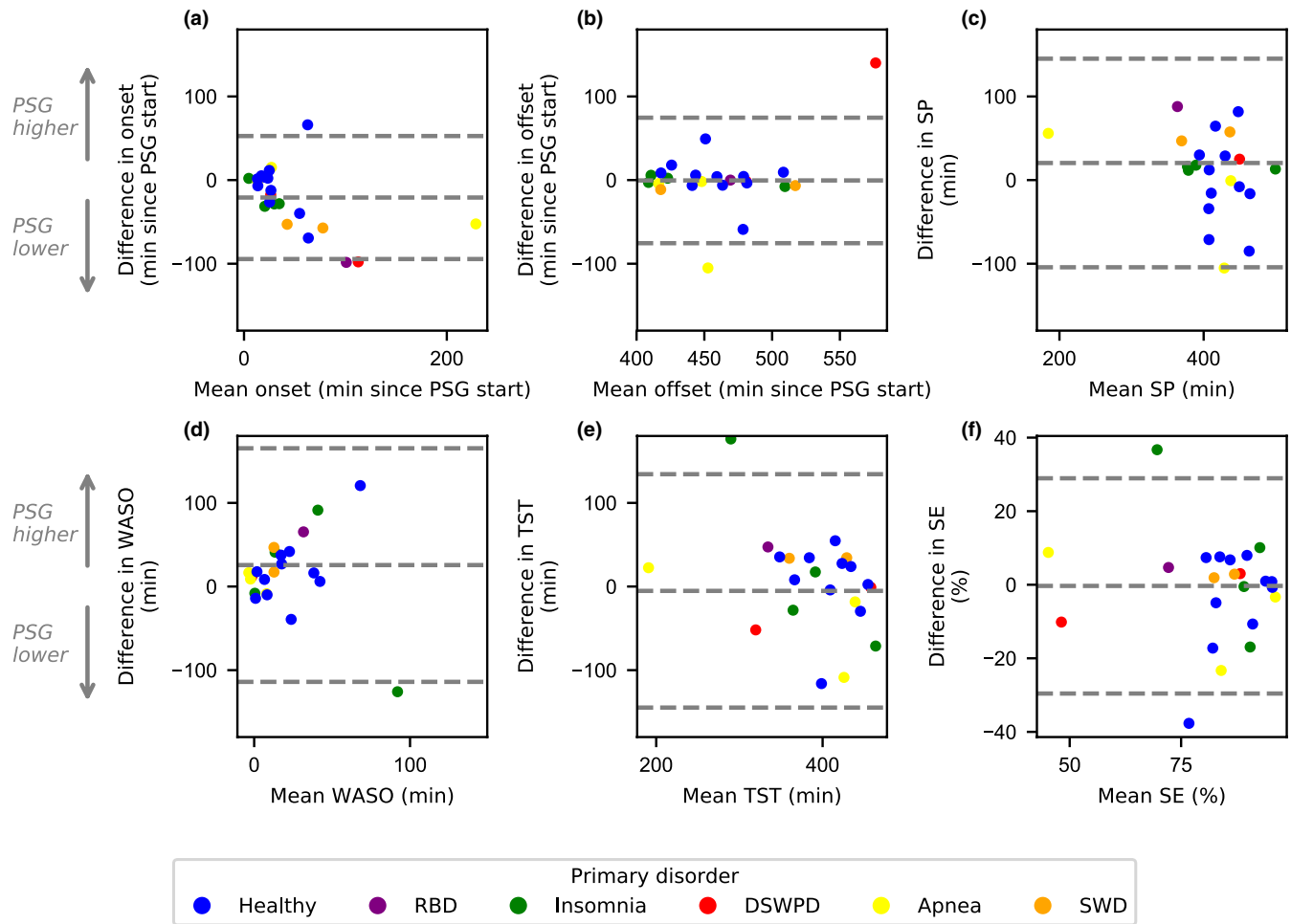
LOOCK ET AL.

Journal of
Sleep
Research

ESRS

9 of 12

FIGURE 5 Bland-Altman analysis of summary sleep parameters from the Munich Actimetry Sleep Detection Algorithm (MASDA) and polysomnography. Bland-Altman plots for (a) sleep onset, (b) sleep offset, (c) sleep period (SP), (d) wake after sleep onset (WASO), (e) total sleep time (TST) and (f) sleep efficiency (SE) of the polysomnography (PSG) sample ($n = 23$). In each panel, the mean between PSG and MASDA is plotted on the x-axis and the absolute difference (PSG-MASDA) on the y-axis. Negative values indicate lower values in PSG than MASDA (for times, this means earlier). The mean difference is denoted by the dashed line in the middle, and the upper and lower boundaries of the 95% confidence interval by the upper and lower dashed lines. Colours indicate the primary sleep disorder diagnosed after the recording night. DWSPD, delayed sleep-wake phase disorder; RBD, rapid eye movement behaviour disorder; SWD, shift work disorder

assessment of sleep onset, WASO, and sleep period, whereas sleep offset, TST and SE were not systematically different between the two methods.

The good agreement of the MASDA and sleep logs in epoch-by-epoch assessments supports previous findings suggesting a reasonable validity between actimetry and sleep logs (Iwasaki et al., 2010; Santisteban et al., 2018; Usui et al., 1999). Validation of MASDA against PSG was also in line with other validation studies, both in terms of actual performance values as well as the overestimation of sleep and underestimation of wake in nocturnal recordings (e.g. Conley et al., 2019; Marino et al., 2013). Specifically, high sensitivity and overall accuracy rates have been reported in numerous validation studies (e.g. Marino et al., 2013; Van de Water et al., 2011). Poor specificity has frequently been reported as a problem of actimetry as well (Dick et al., 2010) and is thus only partly due to the algorithm's design favouring consolidated stretches of sleep. Notably,

we observed specificity and NPV rates ranging from extremely poor performance to perfect concordance with the PSG ratings.

Our results suggest that the comparison of actimetry to a night of PSG is not necessarily appropriate to evaluate the method's 24-h performance. The poor specificity rates obtained in the validation of various actimetry sleep-detection algorithms have often been brought up as a weakness of the method. However, as we show here, actimetry is not by definition worse at detecting wake states. In the sleep-log samples, which include both daytime and night-time data across many days, our algorithm demonstrated good performance in detecting both consolidated sleep (i.e., good sensitivity/PPV) and consolidated wake periods (i.e., excellent specificity/NPV). The low specificity/NPV rates in the PSG sample likely result from the analysis of only about 8 h that are almost exclusively spent in bed, containing mainly sleep epochs, and very few wake epochs, where the wake epochs are marked by little

activity. Indeed, if we assume the plausible scenario that all individuals from the PSG sample were continuously awake 3 h prior to the start of the PSG recording and add these 3 h of wake to the actimetry and PSG records, the rates for sensitivity and PPV remain the same, but median specificity increases drastically from 33% to 86% (IQR = 81%–99%) and NPV from 37% to 86% (77%–100%). Although we cannot be sure that all additional wake epochs would have been recognized as such by the MASDA, this thought experiment illustrates the inherent bias towards low wake-detection if only nocturnal recordings are analysed. Poor specificity values obtained under such conditions may accurately represent the method's difficulty in identifying brief wake interruptions during sleep but not the method's ability to identify longer wake episodes marked by more activity, as occur during the day.

The difficulty of correctly identifying wake interruptions during sleep was also evidenced by the underestimation of WASO by the MASDA. In line with previous research stating that actimetry tends to overestimate sleep and underestimate wake during a sleep episode (e.g., Ancoli-Israel et al., 2003; Van de Water et al., 2011), this finding was particularly expected considering the 24-h moving-average threshold employed by the MASDA. The threshold heavily depends on daytime activity levels and thus trades the underestimation of short sleep interruptions for a high sensitivity for consolidated sleep–wake classification (Roenneberg et al., 2015). In addition, the MASDA's design favouring consolidated stretches of sleep is also likely to contribute to WASO underestimation. In line with Marino et al. (2013), we also noted a systematic increase in this underestimation with a longer average WASO.

Likewise, the observed delay in sleep onset classification by the MASDA in comparison to PSG and the advance in comparison to sleep logs is likely not random. Tryon (2004) introduced the idea of systematic differences between onset scorings because sleep onset has to be understood as a gradual change from wake to sleep. Actimetry typically marks the beginning of a sleep period by immobility (three 10–min bins of immobility under the threshold in the MASDA), whereas PSG considers stereotypical changes in the electrical brain activity measured at the scalp, which can occur later (Marino et al., 2013; Tryon, 2004) or earlier; in log data, it is the subjective perception and recall quality that determines onset and offset times. Whether the opposite findings for the MASDA sleep onsets in comparison to sleep logs and PSG originates only from systematic differences between the three methods or also from sample differences cannot be concluded from our study.

Several limitations have to be put forward in interpreting our results. First, the PSG assessment was conducted in a laboratory environment, which can influence individual sleeping patterns. This setting also called for clearly defined in-bed intervals, which can limit the generalizability of the validation results (Grandner & Rosenberger, 2019), especially the high agreement in regard to sleep offset. Second, the validation was performed in rather homogeneous samples, so the results may not generalize to other populations, particularly people who move very little during the day (bed-ridden or

elderly people). The sleep-log validation was carried out in young, likely healthy sleepers, whereas the PSG validation was performed in a clinical sample where >50% of participants were diagnosed with sleep disorders. Unfortunately, the PSG sample was diminished from 50 to 23 individuals due to software and synchronization issues, and hence we could not analyse effects of particular sleep disorders on the algorithm's performance. Third, our analyses were performed on a resolution of 10 min, so each analysis epoch was only labelled with the most abundant state (sleep or wake) from the PSG 30-s epochs underlying it. This removed information about the relative proportion of sleep and wake within each analysis epoch, precluding the differentiation of performance between clear epochs and "swing epochs". The 10-min filtering that we generally apply to our human activity analyses has proven valuable when long-term, in-context measures of daily sleep–wake behaviour are investigated in contrast to the high-resolution architecture of single nights. Fourth, our diary did not enquire about daytime naps or night-time awakenings. We can only speculate how this might have influenced the MASDA's performance in our sample. MASDA can in principle detect naps, so we would assume that nap information from diaries would have improved the MASDA's performance unless the majority of naps were shorter than the 30-min minimum required by the MASDA. We also assume that information on night-time awakenings could have influenced the MASDA's performance - negatively or positively depending on the type of awakenings reported. Information on short awakenings that are unlikely to be picked up by the MASDA might have reduced performance; long periods with significant tossing and turning or visits to the toilet, which are more likely to be picked up by the MASDA, might have improved performance.

In conclusion, whilst PSG is undeniably richer in detail and more sensitive than the mere monitoring of body movements (Pollak et al., 2001), actimetry can be seen as an objective method of estimating sleep–wake patterns outside the laboratory (Meltzer et al., 2012), supporting large-scale, population-level sleep research. Polysomnography and actimetry are suited for very different questions and settings, and thus are not in competition but should be seen as complementary to each other. By monitoring sleep longitudinally in natural settings, actimetry can help to detect sleep phase alterations, and may assist in the diagnosis of circadian rhythm disorders (Ancoli-Israel et al., 2003; Smith et al., 2018) or the discovery of altered sleep patterns in individuals with sleep or neurobehavioural disorders (Sadeh, 2011). It can also provide objective data on treatment effects of non-pharmacologic and pharmacologic interventions (Ancoli-Israel et al., 2003; Brooks et al., 1993; Roenneberg et al., 2015; Sadeh, 2011; Tryon, 2004). Especially when PSG measurements or sleep logs cannot be obtained, actimetry can contribute greatly to the understanding of individual sleep–wake patterns (Ancoli-Israel et al., 2003; Sadeh, 2011; Smith et al., 2018). We even use it to extract coarse patterns of sleep physiology from wrist movements to assess NREM–REM cycles in the field (Winnebeck et al., 2018). Actimetry is thus a utile tool to measure sleep in diverse populations if conducted using validated algorithms (Smith et al., 2018).

With our sleep detection algorithm, we hope to provide an additional useful and valid tool for studying sleep in the field. In our samples, the MASDA's validity for sleep–wake scoring was in the same range as that for most other algorithms; how it performs in direct comparisons in various study populations needs to be determined in dedicated future studies. What sets the MASDA apart from most other algorithms, is its design to prioritize the detection of consolidated stretches of sleep over detecting frequent changes in the sleep–wake state. Although this is by definition a disadvantage for detailed monitoring of WASO or sleep fragmentation, it is advantageous for circadian analyses striving to assess sleep timing and regularity. There are very few other algorithms with similar designs, but they operate on other principles (e.g., Crespo's or HDCZA in GGIR) (Crespo et al., 2012; van Hees et al., 2018). Importantly, the MASDA is free of assumptions relating to timing, duration and number of sleep bouts per day, so by design, the MASDA is suited to detect sleep at any time of day as often as it occurs (within the limits of its minimum-duration criterion of 30 min). Therefore, the MASDA lends itself particularly to studying shift-working populations, populations with circadian disruption such as jetlag or Non-24, or those with multiphasic sleep.

CONFLICT OF INTEREST

TR, who developed the MASDA, uses the MASDA in the context of the consulting work carried out in Chronsulting UG, of which he is the founder and CSO. All data were collected before TR started active work with Chronsulting. However, to avoid any potential construction of a conflict of interest, TR was not directly involved in this study's data analysis or interpretation. He only provided general guidance and editorial input to the manuscript. In 2020, TR consulted for the Estee Lauder Company, Vanda Pharmaceuticals, Jetlite, Salzgitter AG, NeuroCare, the Vindex Company, KGK Science Ontario, and PricewaterhouseCoopers. All other authors declare that they have no potential conflicts of interest. During the conduct of this study, ASL received a stipend from the Max-Weber Programm (Studienstiftung), NG received research funding from the FoeFoLe program at LMU, LKP received a fellowship from the Coordenação de Aperfeiçoamento Pessoal de Nível Superior-Brasil (CAPES), AMB received travel funds from the ESRS and the Graduate School of Systemic Neurosciences Munich, and CR received a stipend from the Fundação para a Ciência e Tecnologia (FCT) PhD research grants. During the conduct of this study but outside of the submitted work, ECW received travel and research support from the Friedrich-Baur-Stiftung, LMU Excellence, LMU Equal Opportunity, Gordon Research Conference and German Dalton Society.

AUTHOR CONTRIBUTIONS

Conceptualization: TR and ECW. Data curation: AKS, NG, LKP, CM, AMB, MVB and ECW. Formal analysis: ASL. Investigation: CR (PSG), TP (PSG), AKS, NG, LKP, CM, AMB, MVB and ECW. Software: TR (MASDA). Supervision: ECW. Visualization: ASL and ECW. Writing–original draft: ASL and ECW. Writing–review and editing: AKS, CR, NG, LKP, AMB, ASL, TR and ECW.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Catia Reis https://orcid.org/0000-0001-6585-3993
Neda Ghotbi https://orcid.org/0000-0003-3405-4608
Luisa K. Pilz https://orcid.org/0000-0001-7328-6204
Till Roenneberg https://orcid.org/0000-0003-2939-0332
Eva C. Winnebeck https://orcid.org/0000-0002-0717-9432

REFERENCES

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26, 342–392. https://doi.org/10.1093/sleep/26.3.342

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., & for the American Academy of Sleep Medicine. (2016). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: American Academy of Sleep Medicine.e.

Borbély, A. A. (1986). New techniques for the analysis of the human sleep-wake cycle. *Brain Development*, 8, 482–488. https://doi.org/10.1016/S0387-7604(86)80073-0

Borbély, A. A., Rusterholz, T., & Achermann, P. (2017). Three decades of continuous wrist-activity recording: Analysis of sleep duration. *Journal of Sleep Research*, 26, 188–194. https://doi.org/10.1111/jsr.12492

Brooks, J. O. III, Friedman, L., Bliwise, D. L., & Yesavage, J. A. (1993). Use of the wrist actigraph to study insomnia in older adults. *Sleep*, 16, 151–155. https://doi.org/10.1093/sleep/16.2.151

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep*, 15, 461–469. https://doi.org/10.1093/sleep/15.5.461

Conley, S., Knies, A., Batten, J., Ash, G., Miner, B., Hwang, Y., Jeon, S., & Redeker, N. S. (2019). Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 46, 151–160. https://doi.org/10.1016/j.smrv.2019.05.001

Crespo, C., Aboy, M., Fernández, J. R., & Mojón, A. (2012). Automatic identification of activity–rest periods based on actigraphy. *Medical & Biological Engineering & Computing*, 50, 329–340. https://doi.org/10.1007/s11517-012-0875-y

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

Dowle, M., & Srinivasan, A. (2020). *Data.table: Extension of "data.frame."*. https://cran.r-project.org/package=data.table

Ghotbi, N., Pilz, L. K., Winnebeck, E. C., Vetter, C., Zerbini, G., Lenssen, D., Frighetto, G., Salamanca, M., Costa, R., Montagnese, S., & Roenneberg, T. (2020). The μMCTQ: An ultra-short version of the Munich chronotype questionnaire. *Journal of Biological Rhythms*, 35, 98–110. https://doi.org/10.1177/0748730419886986

Girschik, J., Fritschi, L., Heyworth, J., & Waters, F. (2012). Validation of self-reported sleep against actigraphy. *Journal of Epidemiology*, 22, 462–468. https://doi.org/10.2188/jea.JE20120012

Grandner, M. A., & Rosenberger, M. E. (2019). Actigraphic sleep tracking and wearables: Historical context, scientific applications and guidelines, limitations, and considerations for commercial sleep devices. In

M. A. Grandner (Ed.), *Sleep and Health* (pp. 147–157). Cambridge, MA: Academic Press.

Hammad, G., Reyt, M., Beliy, N., Baillet, M., Deantoni, M., Lesoinne, A., Muto, V., & Schmidt, C. (2020). pyActigraphy: open-source python package for actigraphy data visualisation and analysis. *bioRxiv*. https://doi.org/10.1101/2020.12.03.400226

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*, 90–95. https://doi.org/10.5281/zenodo.1343133

Iwasaki, M., Iwata, S., Iemura, A., Yamashita, N., Tomino, Y., Anme, T., Yamagata, Z., Iwata, O., & Matsuishi, T. (2010). Utility of subjective sleep assessment tools for healthy preschool children: A comparative study between sleep logs, questionnaires, and actigraphy. *Journal of Epidemiology*, *20*, 143–149. https://doi.org/10.2188/jea.JE20090054

Kantermann, T., Juda, M., Merrow, M., & Roenneberg, T. (2007). The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Current Biology*, *17*, 1996–2000. https://doi.org/10.1016/j.cub.2007.10.025

Kripke, D. F., Hahn, E. K., Grizas, A. P., Wadiak, K. H., Loving, R. T., Poceta, J. S., Shadan, F. F., Cronin, J. W., & Kline, L. E. (2010). Wrist actigraphic scoring for sleep laboratory patients: Algorithm development. *Journal of Sleep Research*, *19*, 612–619. https://doi.org/10.1111/j.1365-2869.2010.00835.x

Kripke, D. F., Mullaney, D. J., Messin, S., & Wyborney, V. G. (1978). Wrist actigraphic measures of sleep and rhythms. *Electroencephalography and Clinical Neurophysiology*, *44*, 674–676. https://doi.org/10.1016/0013-4694(78)90133-5

Lockley, S. W., Skene, D. J., & Arendt, J. (1999). Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. *Journal of Sleep Research*, *8*, 175–183. https://doi.org/10.1046/j.1365-2869.1999.00155.x

Marino, M., Li, Y. I., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M. (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, *36*, 1747–1755. https://doi.org/10.5665/sleep.3142

Meltzer, L. J., Montgomery-Downs, H. E., Insana, S. P., & Walsh, C. M. (2012). Use of actigraphy for assessment in pediatric sleep research. *Sleep Medicine Reviews*, *16*, 463–475. https://doi.org/10.1016/j.smrv.2011.10.002

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

Pollak, C. P., Tryon, W. W., Nagaraja, H., & Dzwonczyk, R. (2001). How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep*, *24*, 957–965. https://doi.org/10.1093/sleep/24.8.957

Python Software Foundation (2001–2019). *Python*. Language Reference, version 2.7. Available at http://www.python.org.

R Core Team (2020). *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. https://cran.r-project.org/package=psych

Roenneberg, F. T., Keller, L. K., Fischer, D., Matera, J. L., Vetter, C., & Winnebeck, E. C. (2015). Human activity and rest in situ. In A. Sehgal (Ed.), *Methods in enzymology*, Vol. *552* (pp. 257–283). Burlington: Academic Press.

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep Medicine Reviews*, *15*, 259–267. https://doi.org/10.1016/j.smrv.2010.10.001

Sadeh, A., Sharkey, M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep*, *17*, 201–207. https://doi.org/10.1093/sleep/17.3.201

Santisteban, J. A., Brown, T. G., & Gruber, R. (2018). Association between the Munich chronotype questionnaire and wrist actigraphy. *Sleep Disorders*, *2018*, 5646848. https://doi.org/10.1155/2018/5646848

Smith, M. T., McCrae, C. S., Cheung, J., Martin, J. L., Harrod, C. G., Heald, J. L., & Carden, K. A. (2018). Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An American academy of sleep medicine systematic review, meta-analysis, and GRADE assessment. *Journal of Clinical Sleep Medicine*, *14*, 1209–1230. https://doi.org/10.5664/jcsm.7228

te Lindert, B. H. W., & Van Someren, E. J. W. (2013). Sleep estimates using microelectromechanical systems (MEMS). *Sleep*, *36*, 781–789. https://doi.org/10.5665/sleep.2648

Toon, E., Davey, M. J., Hollis, S. L., Nixon, G. M., Horne, R. S. C., & Biggs, S. N. (2016). Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. *Journal of Clinical Sleep Medicine*, *12*, 343–350. https://doi.org/10.5664/jcsm.5580

Tryon, W. W. (2004). Issues of validity in actigraphic sleep assessment. *Sleep*, *27*, 158–165. https://doi.org/10.1093/sleep/27.1.158

Usui, A., Ishizuka, Y., Obinata, I., Okado, T., Fukuzawa, H., & Kanba, S. (1999). Validity of sleep log compared with actigraphic sleep–wake state II. *Psychiatry and Clinical Neurosciences*, *53*, 183–184. https://doi.org/10.1046/j.1440-1819.1999.00529.x

Van de Water, A. T. M., Holmes, A., & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography – a systematic review. *Journal of Sleep Research*, *20*, 183–200. https://doi.org/10.1111/j.1365-2869.2009.00814.x

van Hees, V. T., Sabia, S., Jones, S. E., Wood, A. R., Anderson, K. N., Kivimäki, M., Frayling, T. M., Pack, A. I., Bucan, M., Trenell, M. I., Mazzotti, D. R., Gehrman, P. R., Singh-Manoux, B. A., & Weedon, M. N. (2018). Estimating sleep parameters using an accelerometer without sleep diary. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-31266-z

Vetter, C., Fischer, D., Matera, J. L., & Roenneberg, T. (2015). Aligning work and circadian time in shift workers improves sleep and reduces circadian disruption. *Current Biology*, *25*, 907–911. https://doi.org/10.1016/j.cub.2015.01.064

Webster, J. B., Kripke, D. F., Messin, S., Mullaney, D. J., & Wyborney, G. (1982). An activity-based sleep monitor system for ambulatory use. *Sleep*, *5*, 389–399. https://doi.org/10.1093/sleep/5.4.389

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, 1686. https://doi.org/10.21105/joss.01686

Winnebeck, E. C., Fischer, D., Leise, T., & Roenneberg, T. (2018). Dynamics and ultradian structure of human sleep in real life. *Current Biology*, *28*, 49–59. https://doi.org/10.1016/j.cub.2017.11.063

Winnebeck, E. C., Vuori-Brodowski, M. T., Biller, A. M., Molenda, C., Fischer, D., Zerbini, G., & Roenneberg, T. (2020). Later school start times in a flexible system improve teenage sleep. *Sleep*, *43*, zsz307. https://doi.org/10.1093/sleep/zsz307

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.