Sleep Medicine 85 (2021) 221–229

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

The development, inter-rater agreement and performance of a hierarchical procedure for setting the rest-interval in actigraphy data

Hanna Størksen Follesø [a, 1], Sigrun Borgen Austad [a, 1], Alexander Olsen [a, b], Ingvild Saksvik-Lehouillier [a, *]

[a] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway
[b] St. Olavs Hospital, Trondheim University Hospital, Department of Physical Medicine and Rehabiliation, Trondheim, Norway

ARTICLE INFO

Article history:
Received 12 February 2021
Received in revised form
24 June 2021
Accepted 13 July 2021
Available online 22 July 2021

Keywords:
Actigraphy
Normal sleep adult
Circadian rhythms

ABSTRACT

The aim of this study was to develop and empirically test a hierarchical procedure for defining rest intervals in actigraphy data.

Background: This is a two-part study. The aim of study 1 was to identify common practices for setting rest intervals in actigraphy research and investigate whether standardized guidelines for setting the rest interval exist, as a base to develop a new procedure for defining rest intervals in actigraphy. The aim of study 2 was to empirically test this procedure (The Rest Interval Setting, RISE Procedure). The RISE procedure was applied to a dataset of 537 nights from the sleep study SLEEPIC.

Participants: Participants ($N = 55$) were aged 19–33 ($M = 22.7$, $SD = 3.0$).

Methods: Study 1: Structured overview of the methods used to correct actigraphy data. Study 2: Three scorers independently applied the RISE procedure to the dataset.

Results: Study 1 demonstrated that methods and reporting practices are inconsistent and that there is a need for a standardized procedure for setting the rest interval. The results in study 2 revealed that using the new procedure for setting rest intervals provided high agreement between scorers for both rest onsets ($\alpha = 0.975$) and offsets ($\alpha = 0.998$). Applying the procedure to the dataset resulted in a shortening of the rest interval by 36 min and 19 s on average. There were significant changes ($p < 0.001$) in all sleep estimate outcomes after applying the RISE procedure.

Conclusion: Methods for processing and reporting actigraphy data are highly inconsistent across studies. Here we present empirical support for a new standardized procedure for setting the rest interval, which is likely to improve transparency and reproducibility in achigraphy research.

1. Introduction

Actigraphy is an objective method for sleep assessment. Actigraphy is widely used and allows for continuous, non-invasive longitudinal sleep assessment in large samples at low cost [1]. Standardized guidelines in actigraphy research have repeatedly been requested [2,3] to allow for improved between-study comparability. One aspect of actigraphy research that require standardization is the pre-processing of actigraphic data, specifically the practice of setting the rest interval. To the best of our knowledge, no such common practice exists in the sleep research literature today.

An actigraph is a portable, lightweight, and battery-operated wristwatch device. Several different types of actiwatches exists on the marked, all of them record movements in the form of activity counts through an inbuilt accelerometer [4]. Activity data is typically analysed using software with automatic algorithms tailored for the data collected through the specific actigraph. For each time-epoch, these algorithms integrate activity data from surrounding epochs to the current epoch, scoring it as *sleep* or *wake*. Several sleep variables that can be extracted from actigraphy. *Rest interval duration* is the number of minutes between the time participants begin attempting to fall asleep and the time they get out of bed in the morning or are fully awake. *Total sleep time* is the number of minutes scored as sleep during the rest interval, and *sleep onset latency* is the time between rest onset and sleep onset. *Sleep*

* Corresponding author.
E-mail address: ingvild.saksvik.lehouillier@ntnu.no (I. Saksvik-Lehouillier).
[1] The authors contributed equally.

https://doi.org/10.1016/j.sleep.2021.07.025

efficiency is the percentage of the time in bed spent asleep. *Wake after sleep onset* is the time in minutes between sleep onset and sleep offset scored as wake. The actigraphy software produce an actogram, visualising the sleep data.

Several specific algorithms have been compared [5,6] and validated against polysomnography. Examples of well-established and extensively applied algorithms are the Cole-Kripke algorithm [7,8], the Sadeh algorithm [9], the SDSU algorithm [10,11], and the algorithm validated by Refs. [1,12,13]. An important feature of the different algorithms is that they identify rest intervals; persistently and sufficiently low activity counts where it can be assumed that the participant is resting. Sleep analyses are performed within these rest intervals. It is therefore crucial that the rest interval is defined in an accurate manner. The difficulty in distinguishing between sedentary behavior, rest and sleep [14,15] is a considerable challenge, and it is often necessary to corroborate other inputs for rest onset into the actigraphic data for improved accuracy. Such inputs may be a sleep diary or an event marker. Event markers are buttons the participants can press to mark specific events in time, such as removing the actigraph from the wrist, turning lights off to try to sleep or waking up in the morning. A mark appears in the actogram at the point of time the button was pushed.

Automatically provided rest intervals can be manually reconsidered and set in the software by a scorer. However, there is no consensus regarding how the rest interval should be defined [16]. found that setting the rest interval based on visually inspecting the actigraphic data is superior to the automatic algorithms in setting rest intervals controlled with polysomnography. Others argue that a combination of subjective and automatic scoring provides the best results [1,17] found that actigraphy sleep estimates became more similar to polysomnography results when combined with subjective sleep data. Ancoli-Israel et al. [2] have developed guidelines for actigraphy assessment and recommend having at least one supplemental data source when scoring the rest interval. Self-reported rest onset and offset can validate the algorithmically estimated rest intervals or indicate where they should begin and end. Scientific progress in sleep research is hampered by a lack of consensus regarding setting the rest-interval. Consequently, an empirically informed consensus regarding classification of rest intervals should be given the highest priority by the larger research community.

Here we aimed to develop and empirically test a hierarchical procedure for defining rest intervals in actigraphy data, regardless of the type of actiwatch used. In Study one, we conducted a structured overview of the methods used to explore common practices for setting the rest interval when processing actigraphic data. Based on information from the overview we developed a standardized procedure for setting rest intervals. In Study two, we set out to empirically test the inter-rater agreement and performance of the procedure in a dataset containing a total of 537 nights.

2. Study 1: setting the rest interval in actigraphy: A structured overview

2.1. Method

The basis for our literature search was six major databases Table 1. We included the cross-disciplinary databases Web of Science, SCOPUS, Wiley and Science Direct to ensure the inclusion of articles from a wide spectrum of research fields. Further, Psychnet and PubMed were included. These databases comprise journals in the fields of psychology, medicine and other health sciences. All searches were performed on the same date, April 1, 2018.

The literature search was limited to publications from the past 10 years. The 10-year limit was set to avoid overlap with previous systematic reviews [4,18], and to capture the modern use of actigraphy and possible technological advances.

2.1.1. Exclusion criteria

The total number of research articles that resulted from the literature search was 4260 (Fig. 1). All articles were imported to EndNote X8. After removing duplicates, 2566 articles remained.

a) Articles with non-human samples.
b) Articles with samples including participants younger than age 12. We chose this threshold because children under the age of 12 may not be eligible to complete their own sleep diary or consistently press the event marker [19,20]. Research on this group may benefit from a tailored procedure for setting the rest interval in actigraphy.
c) Studies assessing sleep in Parkinson's disease and restless leg syndrome were removed due to involuntary movements creating a potential need for specialized algorithms and procedures in this population [21].
d) Articles assessing sleep using other instruments than actigraphy (e.g, polysomnography, sensor in mattress or placed in the room, questionnaires, ambulatory blood pressure monitoring, sleep logs).
e) Articles using consumer-targeted activity watches instead of research-targeted accelerometers were removed due to differences from actigraphy in polysomnography validation studies [22,23].
f) Studies where the actigraph was placed on other locations than the wrist (i.e, ankle or waist) were excluded due to the possibility of differences in activity patterns for different placement locations [24].
g) Studies where the authors did not edit the raw actigraphy data because the aim of the study was to compare raw actigraphy data to other sources of sleep data

2.1.2. Extracting information

All remaining articles ($N = 1061$) were examined for material about rest interval setting, algorithm application, the use of event markers and sleep diaries. Furthermore, to ensure the collection of all relevant data, we applied a search within the article comprising the keywords, “algorithm”, “software”, “diary”, “sleep log”, “button”, “event marker”, “marker” and “actigraphy”.

2.1.3. Analyses

For each article, we noted whether sleep diaries and/or event markers were present. This information was coded into Yes/No conditions. We also noted the reported degree of corroboration of sleep diaries and/or event markers into the actigraphic data, the reported software and algorithms applied and validation study citations for the algorithm. Finally, we determined whether articles reported using technicians or professionals to score the actigraphic data and if intraclass correlations for scorers were reported. We clustered articles that had similar reporting of data processing methods. Information extracted from the articles was imported into IBM SPSS Statistics 25 and performed descriptive analyses.

2.2. Results

In this structured overview over existing literature we found inconsistent practices in actigraphic data processing. We identified clusters of articles describing how researchers report the processing of their actigraphic process; specifically, how they set rest intervals, how they analysed the data and if they collected event markers and/or sleep diaries.

Table 1
Keywords Applied in Literature Search.

	[a]Web of Science	SCOPUS	Wiley	Psychnet	Science Direct	*PubMed
Search criteria	Actigraphy AND Sleep in Topic	Actigraphy in Keywords AND Sleep in Abstract	Actigraphy in Keywords AND Sleep in Abstract	Actigraphy in Keywords AND Sleep in Abstract	Actigraphy in Keywords AND Sleep in Abstract	Actigraphy in MeSH Major Topic AND Sleep in Abstract
Number of hits	1944	1514	148	20	333	301

[a] *Note.* The search engines in Web of Science and PubMed had different search options compared to the other databases. Instead of "keyword" and "abstract" options, these databases offered to search in "topic" and "major topic", and thus have different search specifications. We attempted to perform as similar searches as possible for all search engines.

We found two main types of articles; one in which no additional data sources other than actigraphy were collected, and one in which sleep diary and/or event markers were collected along with actigraphy (see Fig. 2). Practices varied within each group. Articles that collected sleep diaries and/or event markers were further divided into three clusters. The first cluster consisted of articles that did not report any interaction between the sleep diary and/or the event marker and the actigraphic data; the second contained articles that reported corroboration of sleep diaries and/or event markers into the actigraphic dataset, but not how this corroboration was performed in detail. The third cluster was composed of articles that either reported or cited a stepwise procedure for setting the rest interval.

A total of 252 research articles (23.75%) from our literature overview did not report collecting sleep diaries nor event markers (Fig. 2). In all, 809 research articles (76.25%) from our literature overview reported collecting sleep diaries and/or event markers. We divided these articles into clusters. In the first cluster, there are 449 articles, all of which collected sleep diaries and/or event markers but presented no information about how these inputs were used in the actigraphic scoring. The second cluster consists of 349 articles, which describe a corroboration of the sleep diary or event marker data with software analyses. The last cluster of 11 articles provided or cited a stepwise procedure for setting the rest interval. One of the articles in this group provided a validated, inter-rater tested procedure for setting the rest interval [17]. Furthermore, the remaining ten articles on this cluster either referred to a validated procedure providing interscorer correlations for outcome variables, but not for rest intervals [25] or provided stepwise and detailed explanations of how they integrated sleep diaries and/or event markers with their actigraphic data, but that procedure is not reported as validated or inter-rater tested.

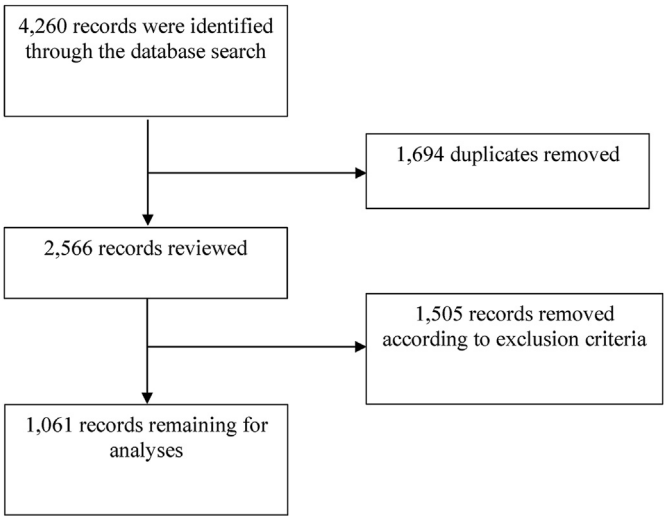

Fig. 1. Number of records included and excluded in each step of the study selection.

2.3. Discussion

In this study, we examined how the processing of actigraphic data was reported, how rest intervals were set and how the data was analysed in 1061 actigraphy studies. We found large variations in the practice of data processing. We found no established procedure that governed common practice for actigraphic data processing.

A major observation is that most actigraphic studies collect sleep diaries and/or event markers simultaneously with the actigraphic data (809 out of 1061 in our overview, 79.63%). This is in concordance with actigraphy data scoring recommendations by Ref. [2] where it is recommended that scorers have at least one supplemental source of data in addition to the actigraphic data when setting the rest interval. However, the number of studies that collected event markers (232 studies, 23.83%) was considerably lower than the number of studies that collected sleep diaries (767 studies, 75.79%). This implies that the available, or preferred, supplemental data source for actigraphy over the past 10 years has been the sleep diary.

There was great variation in data processing reporting practices. Many studies cited a validated algorithm for actigraphic data analyses, but a majority did not mention which specific algorithm was applied. This also held true for software reporting, as many papers reported that their analyses were conducted using a specific software, but a majority of these did not specify the brand or version of the software. Furthermore, there were large variations in the detailing of the descriptions of how the rest intervals were set, including the report of corroboration of supplemental sources of data. The inconsistencies that appeared in our overwiev are in line with a previous literature review conducted more than 10 years ago [18] that found no common practice for reporting actigraphic data processes.

2.4. Conclusion

In our overview, we found that researchers report processing actigraphic data differently, and thereby confirm that there is a need for an accepted and commonly used way of setting rest intervals. This raises the question of comparability and reproducibility across actigraphy studies. We did find procedures that were inter-rater tested, but these were rarely cited in the articles included in our overview and thus cannot be said to have gained traction in actigraphy research. To address this issue, we propose a hiearachical procedure, the Rest Interval Setting (RISE) Procedure, for setting rest intervals based on the existing literature.

3. Development of the RISE procedure

We identified four stepwise procedures for setting the rest intervals. One provide an inter-rater tested, stepwise procedure for setting the rest interval [17] relying on light, activity and event markers. Another procedure [25] suggest varying thresholds of

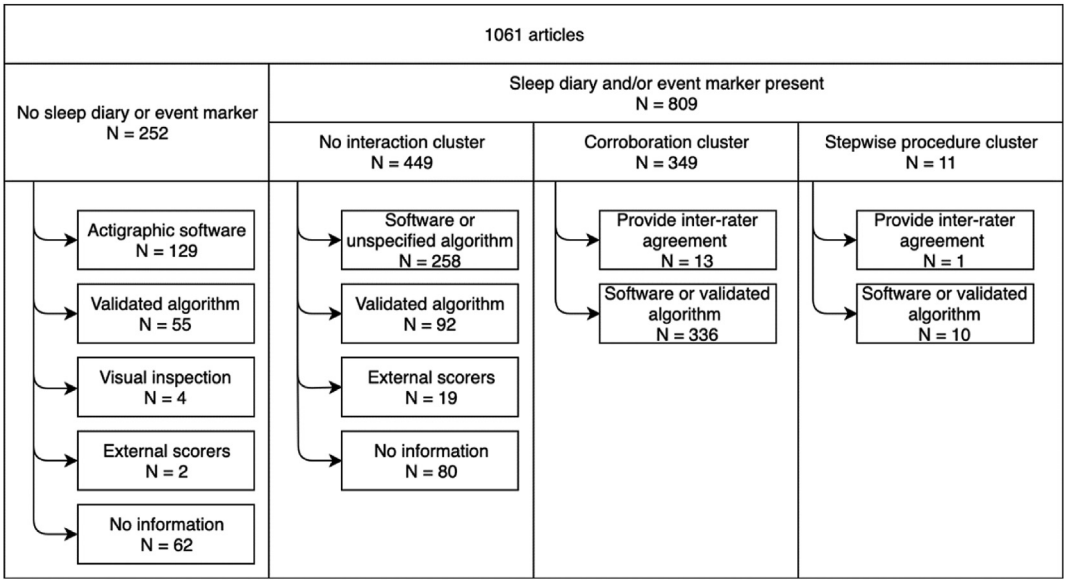

Fig. 2. Overview of clusters. The two main partitions and the three clusters of interaction between sleep diaries and/or event markers and actigraphic data are illustrated.

accepted agreement between sleep diaries, event markers, activity and light data. Later studies with the same author have cited this procedure [26]. assessed five different approaches for setting the rest intervals.They assessed the use of sleep diary, event marker, automatic algorithm and visual inspection to set the rest interval. They also assessed the corroboration of the automatic algorithm with an event marker. The different approaches were compared with polysomnography. This study was not a stepwise procedure for setting rest intervals, but it provides information about the strengths and weaknesses of each approach. Finally, the MESA study [27] developed a stepwise procedure for setting the rest interval, assessing event marker, sleep diary, activity and light. No scientific publication has tested the reliability of this procedure.

Procedures for actigraphic data editing apply different thresholds of accepted agreement between data sources. Agreement within a threshold refers to the accepted distance in time between one input and another. Thresholds ranged from 10 min [17,28], to 15 min [25,27] to 30 min [29]. [25] applied both a 15 min and 30 min threshold. [17] assessed their procedure by investigating their inter-scorer agreement and successfully scored rest intervals. Only two inputs were assessed at the same time in this study. In our procedure, we sought to set the rest interval by initially assessing inputs three by three (rest interval set by algorithm in the software, sleep diary and event marker). Thus, we chose a threshold of 15 min and required that at least two of three inputs fall within 15 min of each other for the rest interval to be considered valid. When using three inputs we considered a 10 min threshold to be too stringent, as this could result in the exclusion of a large portion of the data. Still a 30 min threshold may be too long. The rationale behind this argument is that the longer the threshold is, the risk of accepting rest intervals with measurement errors in either one or more inputs increases. To the best of our knowledge, no procedures using a threshold has been validated against PSG. Thus, until a threshold is empirically established we suggest a 15 min threshold.

The RISE procedure follows the example of previous research granting the event marker precedence over the sleep diary [30,31]. The event marker is argued to be less prone to human error than the sleep diary [32] since participants are not required to remember what time they pressed the button the morning after, as they do with diaries. Use of the event marker is, however, not without fault, as participants may forget to press the button, resulting in missing data. The validity of actigraphy increases when it is analysed in conjunction with sleep diaries [1], therefore we include both event markers and sleep diaries in the RISE procedure. With the event marker being a flexible tool for data collection, it is important that participants in studies planning to apply the RISE procedure are well instructed to press the event marker to define rest on- and offset. The event marker can, as mentioned, define other events such as removing the watch or participants can be instructed to press the button as close to falling asleep as possible. For the RISE procedure we wish to identify rest with the intention to sleep, assisting researchers to differentiate idle behavior from sleep behaviour.

Light input is included in the RISE procedure with caution, as natural light may vary across seasons, and be out of the individuals control. In such cases, changes in light would not be indicative of sleep behaviour. Despite the vulnerability of the light condition, it may be helpful. For example, Chow and colleagues [17] successfully used the light variable as a help criterion and obtained a high inter-rater reliability (Cronbach's alpha was 0.975 for bedtime and 0.995 for rise time) in defining the rest interval.

A visualisation of the procedure can be found in Fig. 3.

Prior to applying the RISE procedure to the actigraphic data, researchers will need to run an initial sleep analysis in the software to create default rest intervals. These default rest intervals will be confirmed, edited or excluded by applying the procedure. The first step of the proedure (Fig. 3) is the automatic *screening* process, where rest intervals that do not require editing are identified. A rest interval does not require visual editing when the compared inputs cited in Fig. 3 occur within 15 min of each other. The screening process starts in zone A, where entries of event markers, sleep diaries and the algorithmically defined rest intervals are compared. If the event marker or sleep diary is missing, the screening process takes place respectively in zone B, comparing event marker and algorithmically defined rest interval, or in zone C, comparing sleep diary and algorithmically defined rest interval.

Intervals that are not validated in the screening process need to be manually assessed in order to set the rest interval. The first zone for

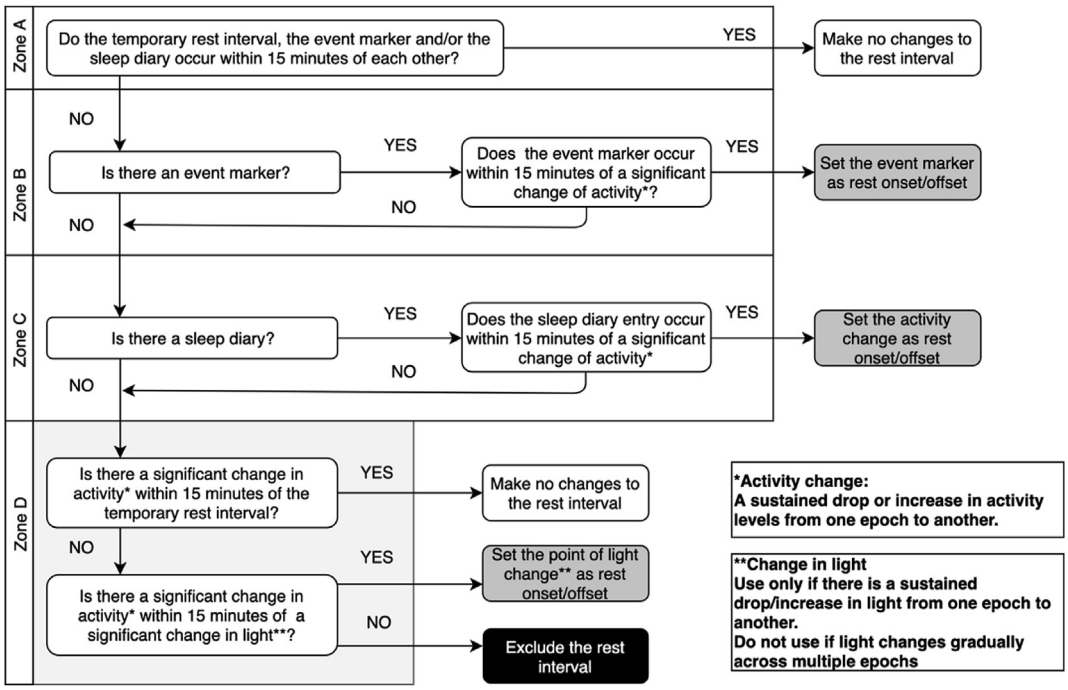

Fig. 3. Procedure for setting the rest intervals in actigraphic data.

manual assessment is zone B. Here, the scorer assesses the event marker entry and corroborates it with a visual inspection of activity decrease/increase. The rest interval onset/offset is set to the event marker entry if it falls within 15 min of a significant change in activity. If the event marker entry occurs outside the 15-min threshold, or the event marker is not present, the sleep diary is assessed in zone C. If the diary entry falls within 15 min of a significant activity change, the point of significant activity change is set as the rest onset/offset.

If neither the event marker entry nor the sleep diary entry occurs within 15 min of activity change, or if the event marker and sleep diary are missing, researchers are restricted to zone D of the manual. In this zone, the initially defined rest intervals from the algorithm are compared to changes in activity and light. Data should be excluded if these do not fall within 15 min of each other. To avoid biases when manually scoring rest intervals using zone D, we applied stipulations to the activity and light level criterion: if there is a sudden drop/increase in activity or light from one epoch to the next, and this change occurs within 15 min of the other variables, the initially defined rest interval is considered validated by activity and/or light. Specifically, for light, if the changes are more gradual across multiple epochs, the light condition is considered invalid and should not be applied [25].

3.1. Study 2 development of a hierarchical procedure for setting the rest-interval in actigraphy data

The objective of this study was to empirically test the reliability of the RISE procedure for editing and setting rest intervals and the effect on sleep outcome variables. Three independent scorers applied the procedure to an actigraphic dataset from a study investigating effects of sleep deprivation in healthy subjects using actigraphy [33].

3.1.1. Participants

A total of 55 healthy subjects, 11 male (20%) and 44 female (80%), were included in the SLEEPIC study (I [34]. Mean age was 22.69 (SD = 3.02, range 19–33). Participants were recruited as a convenience sample through social media, university lectures and information on bulletin boards throughout campus at the Norwegian University of Science and Technology.

3.1.2. Procedure

3.1.2.1. Data collection. Actigraphic data was collected in five waves across 10-day periods. The participants wore an actigraphic wrist watch (Philips Actiwatch Spectrum PRO) on their non-dominant wrist and completed a personal sleep diary [35] for 10 consecutive days. They were instructed to remove the actigraph only before bathing or using a sauna. The participants were asked to press an event button on the actigraph to mark “lights off” when they were going to sleep, and again when they woke up in the morning. All participants received a leaflet with this information. Data was collected in 15-s epochs, with a medium sensitivity threshold of 40 activity counts. The data was analysed by the default algorithm in the Actiware version 6.0.9 software, an algorithm validated by Refs. [1,13] No initial changes were made to the primary actigraphy set to deal with missing data, as this is a part of the proposed RISE procedure (see Fig. 3 eg. there is information of what steps to take if the participant has not used the event marker for one of the days). The procedure does not address issues with missing actigraphic data, and nights where activity data was missing were excluded from the data set.

Participants completed the sleep diary [35] partly in the morning noting what time they went to bed, turned the lights off and the time lapse between lights off and falling asleep the evening before. They also registered wake-up-time, rise time and a rating of the previous night's sleep, number and duration of awakenings, and information about the use of sleep medication and alcohol consumption. In the evening, the participants registered the number of naps and perceived function during the day. Completing the sleep diary was estimated to take 5 min each day.

3.1.2.2. Application of the RISE procedure. Three independent scorers applied the novel RISE procedure for setting the rest interval to the SLEEPIC data set. RISE was developed by two of the three scorers (HSF and SBA). Scorer 1 (AMN) received training from the developers, and then all scorers applied it to the dataset independently. The initial screening process of the procedure was conducted before the scorers applied the procedure using IBM SPSS Statistics 25. The scorers assessed the remaining intervals.

3.1.2.3. Ethical considerations. The study was approved by the Regional Committee for Medical and Health Research Ethics in Central Norway (REK number 2017/85) and was in accordance with the 1964 Helsinki Declaration and its later amendments.

3.1.3. Analyses

All statistical analyses were conducted in IBM SPSS Statistics 25. Initial analyses were performed, including descriptive analyses of the raw actigraphic data and of the screening process by which rest interval on- and offsets that did not require manual inspection were identified.

To assess agreement between scorers we first assessed in which zones scorers assessed each rest interval on- and offset. Thereafter, we assessed intraclass correlation (ICC) according to published guidelines [36], by running a reliability analysis and choosing the intraclass correlation coefficient in SPSS. Because we had three fixed scorers, we chose the two-way mixed ICC model. Pearson correlation between scorers and internal consistency was assessed using Cronbach's alpha.

To assess the significance of inter scorer disagreement; we isolated cases where scorers had produced different rest intervals and conducted a paired sampled *t*-test on these cases. We then compared sleep outcome measures before and after the data had been edited using the RISE procedure, selecting one scorer (HSF). Paired-sample *t*-tests were conducted for this purpose. These analyses were performed for rest interval duration, total sleep time, sleep onset latency, sleep efficiency and wake after sleep onset.

To assess how many rest intervals would be accepted in the screening procedure in a dataset with a full set of event markers and sleep diaries, we extracted a subsample consisting of rest intervals with both event marker and sleep diaries available. We ran descriptive analyses and identified the number of rest intervals that were accepted in the screening process.

3.2. Results

3.2.1. Initial analyses

3.2.1.1. Actigraphic dataset. The raw dataset consisted of 537 actigraphically measured nights, a total of 1074 cases of bedtime and rise time. A total of 15 out of 59 participants were instructed to press the event marker to indicate sleep onset and offset. As a result event markers were present for 143 rest interval onsets (394 missing) and 49 rest interval offsets (488 missing). In this sub sample, during the weekdays the event marker was pressed in 88% of the cases for rest interval onset and 30% for the rest interval offset, while in the weekend 48% pressed the event marker for rest onset, and 16% for rest offset. Sleep diaries were present for all rest interval onsets and offsets.

3.2.1.2. Screening process. For rest *onset*, the screening process automatically identified 199 out of 537 (37.1%) cases, leaving 338 cases to be processed by the scorers. For rest interval *offset*, 343 out of 537 (63.9%) cases were identified in the screening process, leaving 194 cases to be processed by the scorers.

Table 2a
Pearson correlations between scorers for rest onset and offset.

		Scorer 1	Scorer 2	Scorer 3
Rest interval onset	Alpha if item excluded	0.971	0.958	0.962
	Scorer 1	1	0.926	0.920
	Scorer 2		1	0.944
	Scorer 3			1
Rest interval offset	Alpha if item excluded	0.997	0.997	0.998
	Scorer 1	1	0.996	0.994
	Scorer 2		1	0.994
	Scorer 3			1

We conducted the screening process on the subsample consisting of rest on- and offsets where both event markers and sleep diaries were present. The subsample consisted of 148 rest onsets and 39 rest offset cases. For rest *onset*, 79 cases (55.24%) were identified in the screening process leaving 69 cases (45.76%) for inspection. For rest *offset*, 45 cases (91.84%) were identified in the screening process, leaving 4 (8.16%) for inspection.

3.2.2. Agreement between scorers

3.2.2.1. Intraclass correlation. For rest interval *onset*, average measure intraclass correlation (ICC) was 0.976, 95% CI [0.972, 0.979] and the single measure ICC was 0.930, 95% CI [0.920, 0.940] ($F(522,1044) = 41.11$, $p < 0.001$). The average measure ICC for rest interval *offset* was 0.998, 95% CI [0.998, 0.999] and single measure ICC was 0.995, 95% CI [0.994-0.996] ($F(525,1050) = 598.32$, $p < 0.001$). Correlations between scorers two-by-two and alpha if item excluded are reported in Table 2a.

3.2.2.2. Difference between scorers. To assess potential inaccuracies of the RISE proecdure, we isolated the rest interval onsets and offsets where the rest intervals of each scorer differed after applying the procedure. Scorers one and two differed on 172 (32%) rest interval onset cases and 62 (22.55%) rest interval offset cases. Scorer one and three differed in 177 (32.96%) rest interval onset cases and 71 (13.22) rest interval offset cases. Similarly, Scorers two and three had differing responses in 176 (32.77) rest interval offset cases and 75 (13.97%) rest interval offset cases. The differences between scorers are reported in Table 2b.

3.2.3. Outcome measures from one scorer

The results presented below are comparisons of the sleep variables extracted when using the default rest intervals and rest intervals as determined by one of the scorers' (scorer two) after application of the RISE procedure. We chose scorer two because this scorer had the lowest alpha if an item was removed. Prior to applying the procedure, there were 537 actigraphic nights. The total number of nights after data editing was 520. Thus 3.17% of the entire dataset was excluded.

3.2.3.1. Rest intervals. The mean rest interval duration was 08:15:06 (hh:mm:ss) ($SD = 02:16:28$, range = 03:03:45–21:55:15) initially. After applying the procedure, the mean rest interval duration was 07:37:20 ($SD = 01:47:24$, range equals; 03:03:45–13:50:00). The paired-sample *t*-test showed a significant change in rest interval duration, $t(518) = 8.64$, $p < 0.001$ (two-tailed). The mean decrease in rest interval duration was 36 m 19 s after applying the procedure, with a 95% confidence interval ranging from 00:28:04 to 00:44:34.

Out of the 520 rest intervals, Scorer two made changes to 224 rest intervals (43.08%). In 17 cases, the rest interval duration was shortened

Table 2b
Mean Differences Between Scorers When Disagreement Occurred.

	Mean difference	*SD*	95% confidence interval		
			Lower	Upper	*t*-value
Rest interval onset					
Scorer 1 and 2	00:15:36	00:42:58	00:09:07	00:22:04	$t(171) = 4.76$, $p < 0.001$
Scorer 1 and 3	00:14:00	00:46:55	00:07:02	00:20:57	$t(176) = 3.97$, $p < 0.001$
Scorer 2 and 3	00:00:21	00:29:32	−00:04:02	00:04:45	$t(175) = 0.159$, $p = 0.873$
Rest interval offset					
Scorer 1 and 2	−00:00:57	00:25:12	−00:07:21	00:05:26	$t(61) = -0.297$, $p = 0.767$
Scorer 1 and 3	−00:01:30	00:28:48	−00:08:19	00:05:18	$t(70) = -0.443$, $p = 0.659$
Scorer 2 and 3	−00:00:31	00:27:22	−00:06:49	00:05:46	$t(74) = -0.168$, $p = 0.867$

by five or more hours. 16 rest intervals were modified by 3–5 h, by 3–5 h. In addition, 57 rest intervals were modified by 1–3 h. The remainder of the rest intervals (134 intervals) were modified by −1 to 1 h.

3.2.3.2. Variables calculated by the algorithm in the software. Mean total sleep time from the initial analyses providing default rest intervals was 06:42:10 (*SD* = 01:41:14, range: 00:16:00–18:14:00). After applying the procedure, mean TST was 06:36:43 (*SD* = 01:35:09, range 02:37:15–12:15:30). The paired-sample *t*-test showed a significant difference in TST, $t(518) = 3.88$, $p < 0.001$ (two-tailed). The mean decrease in TST after manual correction was 6 m 50 s with a 95% confidence interval ranging from 00:03:22 to 00:10:18.

Sleep onset latency was initially calculated to 34.82 min (*SD* = 54.54, range: 0 m–494.5 m) by the algorithm in the software. In the corrected dataset, sleep onset latency was calculated to 14.73 min (*SD* = 20.02, range: 0 m–174 m). The paired-sample *t*-test showed a significant difference in sleep onset latency, $t(518) = 8.27$, $p < 0.001$ (two-tailed). The mean decrease in sleep onset latency after manual correction was 19.69 min with a 95% confidence interval ranging from 15.02 to 24.38.

Sleep efficiency was initially calculated to 83.00% (*SD* = 10.24%, range: 4.69%–97.57%) by the algorithm. After applying the procedure, sleep efficiency was 86.85% (*SD* = 6.2%, range: 47.14%–97.57%). The paired-sample *t*-test showed a significant difference in sleep efficiency, $t(518) = 9.86$, $p < 0.001$ (two-tailed). The mean increase in sleep efficiency after manual correction was 3.85% minutes with a 95% confidence interval ranging from 3.08% to 4.62%.

Wake after sleep onset was initially calculated to 29.59 min (*SD* = 22.53m, range: 0 m–243.75 m) by the algorithm, shifting to 25.66 min (*SD* = 13.76m, range: 3.5m–92.5m) after applying the procedure. The paired-sample *t*-test showed a significant difference in wake after sleep onset, $t(518) = 4.78$, $p < 0.001$ (two-tailed). Similar to sleep efficiency, the mean decrease in wake after sleep onset after manual correction was 3.63 min with a 95% confidence interval ranging from 2.14 to 5.12.

3.2.3.3. Subsample analysis. A paired-sample *t*-test compared data extracted from the default rest intervals to a subsample of the edited rest intervals, isolating rest intervals that Scorer two had redefined using the RISE procedure (223 rest intervals). The mean decrease in rest interval duration compared to default rest intervals for the subsample was 01:24:49 $t(223) = 9.69$, $p < 0.001$ (two tailed), 95% confidence interval ranging from 01:07:33 to 01:42:04. The mean decrease in total sleep time was 00:15:58, $t(221) = 4.03$, $p < 0.001$ (two tailed), 95% confidence interval ranging from 00:08:10 to 00:23:45. Mean decrease in Wake after sleep onset was 00:08:24, $t(221) = 4.89$, $p < 0.001$ (two tailed), 95% confidence interval ranging from 00:05:01 to 00:11:49. Mean decrease in Sleep onset latency was 00:46:04, $t(221) = 9.10$, $p < 0.001$ (two tailed), 95% confidence interval ranging from 00:36:05 to 00:56:01.

4. Discussion

In this study, we applied a novel procedure for setting rest intervals to a data set of 1074 rest onsets and offsets, the RISE procedure. The purpose was to investigate inter-scorer validity of the procedure, and its effect on sleep outcome variables. We observed high inter-rater agreement, meaning that the RISE procedure provides a homogenous correction of rest intervals. The procedure removed noise in the dataset, and led to statistically significant changes in all sleep estimates.

4.1. Screening process

A substantial portion of the rest intervals was successfully identified as valid in the automatic screening process (37.1% rest onset cases, 63.9% rest offset cases). The number of missing event markers was high. In a sample with a higher amount of present event markers, the screening process would identify an even larger amount of rest intervals. This was demonstrated in the subsample analysis with a full sample of event markers and sleep diaries, where 55.24% of rest onset cases and 91.84% of rest offset cases were identified as valid by the screening process. Thus, by rigorously collecting sleep diary and event markers from participants, a large proportion of the rest intervals will not require visual inspection when using our procedure. There participants seemed to use the event marker more rigorously in the weekdays compared to weekends, future research should emphasize the importance of using the event marker also in the weekends for the participants. However, it is not likely that this affect the use of the RISE procedure or the conclusions of this study.

4.2. Agreement between scorers

Overall, our results indicate high agreement between scorers. Scorers two and three had a slightly higher agreement and lower disagreement compared to scorer one. Agreement with scorer one, however, was nonetheless high, and the interscorer agreement in this study is similar to the results of Chow and colleagues [17] (Cronbach's alpha of 0.975 for bedtime and 0.995 for rise time).

We isolated the cases where scorers disagreed to assess the size and significant of the disagreement. Scorers disagreed upon 32–33% of the rest onset cases and 13–22% of the rest offset cases. In these cases, disagreement for rest offset was not statistically significant. For rest onset, disagreement between scorer 1 and the other scorers was 14 and 15 min and significant. This implies that scorers may interpret some inputs from the software differently, especially for rest onset. This disagreement often occurred where

the activity change input was ambiguous. In these cases, the scorers need to make a subjective evaluation of sustained activity change, and such an evaluation may differ between scorers.

4.3. Outcome measures from one scorer

All sleep variables changed significantly after applying the RISE procedure. The most evident change to the raw actigraphic data was that the delineation of the rest interval in almost all incidents was shortened. This generated shorter rest intervals along with shorter total sleep time compared to the default rest interval. The reduction of the rest interval duration was a consequence of the participants' self-reported rest onset occurring later in the evening than the default rest interval. The RISE procedure contributed to a significant noise reduction in the dataset, which was notable considering the many rest intervals that were changed by several hours. In the subsample consisting of nights where changes were made, the mean changes in rest interval duration were 01:24:49 and 00:15:58 for sleep duration. This implies that the algorithm in the software often is precise (57.1% of the cases), but where corrections are needed, the error of the algorithm is often substantial for rest interval duration when controlled using activity assessment, sleep diaries, event markers and changes in light.

Actigraphy is known to overestimate sleep and underestimate wake [1,14], and applying our proposed procedure shortened the total sleep time compared to the initially estimated sleep time. However, until this procedure is validated against polysomnography, we cannot ascertain whether the observed changes in total sleep time occur where the actigraph misinterprets wake as sleep.

The purpose of the procedure was to improve the boundaries within which the algorithm sleep analysis is conducted. Thus, common errors in actigraphically measured sleep are not ameliorated. Such errors include underestimation of wake within the sleep interval [1]. What is amended for is the miscalculation of rest interval onset and offset and thus false sleep epochs within these intervals.

4.4. Strengths and limitations

In this study, the RISE procedure was tested on a non-clinical sample. Future research should aim to validate this procedure against polysomnography as well as in other specific and relevant samples, such as insomniacs. Furthermore, the procedure was applied by the authors, and thus, other researchers should apply the RISE procedure to their datasets and investigate their agreement. Finally, to ensure the best feasibility and quality of the proposed procedure, it should be tested using other actigraph brands, other software packages and algorithms.

Strengths include using a large sample to assess the trends for how researchers process their actigraphic data. The model is based on several approaches to setting rest intervals, thereby providing a comprehensible method that is also easy for researchers to apply. The screening process in the proposed model will likely reduce time spent on editing invalid rest intervals when using actigraphy for sleep assessment.

5. Conclusion

In sum, the RISE procedure for setting the rest interval was developed as a straightforward and stepwise tool. It has a screening procedure and distinct criteria for exclusion of rest intervals. Using this procedure will provide increased transparency and reproducibility across actigraphy studies. We acknowledge that inputs that require subjective evaluation may increase the risk of between-scorer variation.

When testing the RISE procedure, we achieved high inter-rater agreement ($\alpha = 0.975$ for rest onset, $\alpha = 0.998$ for rest offset), and thus, this procedure provides a homogenous correction of rest intervals. On average, the rest interval was shortened by 36 min and 19 s, there were significant changes in all sleep estimates and applying the procedure removed noise in the dataset. Nevertheless, future studies should test whether the RISE procedure improves the accuracy of actigraphy sleep data in relation to PSG.

A large number of rest intervals were identified in the automatic screening process. However, not all rest intervals will be automatically approved, confirming a need for standardized procedures for manually setting rest intervals even when researchers rigorously collect event markers and sleep diaries, as offered in the procedure. This is currently the only procedure with a standardized hiaerachical screening process that will save researchers time when setting the rest interval compared to existing procedures.

Credit author statement

Hanna Størksen Follesø: Conceptualization, Methodology and analyses, Writing – original draft, Writing – review & editing, Sigrun Borgen Austad: Conceptualization, Methodology and analyses, Writing – original draft, Writing – review & editing, Alexander Olsen: Conceptualization, Methodology and analyses, Writing – review & editing, Project administration, Ingvild Saksvik-Lehouillier: Conceptualization, Methodology and analyses, Writing – review & editing, Project administration

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Acknowledgements

We thank all the participants in this project, the student assistants who contributed to the data collection and the Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU).

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: https://doi.org/10.1016/j.sleep.2021.07.025.

References

[1] Kushida CA, Chang A, Gadkary C, et al. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001;2(5):389–96. https://doi.org/10.1016/S1389-9457(00)00098-8.

[2] Ancoli-Israel S, Martin JL, Blackwell T, et al. The SBSM Guide to actigraphy monitoring: clinical and research applications. Behav Sleep Med 2015;13:4–38. https://doi.org/10.1080/15402002.2015.1046356.

[3] Sadeh A. The role and validity of actigraphy in sleep medicine: An update. Sleep Med Rev 2011;15(4):259–67. https://doi.org/10.1016/j.smrv.2010.10.001.

[4] Ancoli-Israel S, Cole R, Alessi C, et al. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003;26(3):342–92. https://doi.org/10.1093/sleep/26.3.342.

[5] de Souza L, Benedito-Silva AA, Pires MLN, et al. Further validation of actigraphy for sleep studies. Sleep 2003;26(1):81–5. https://doi.org/10.1093/sleep/26.1.81.

[6] Kim MJ, Lee GH, Kim CS, et al. Comparison of three actigraphic algorithms used to evaluate sleep in patients with obstructive sleep apnea. Sleep Breath 2013;17(1):297–304. https://doi.org/10.1007/s11325-012-0689-z.
[7] Cole HV, Kripke DF, Gruen W, et al. Automatic sleep/wake identification from wrist activity. Sleep 1992;15(5):461–9. https://doi.org/10.1093/sleep/15.5.461.
[8] Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013;36(11):1747–55. https://doi.org/10.5665/sleep.3142.
[9] Sadeh A, Sharkey M, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994;17(3):201–7. https://doi.org/10.1093/sleep/17.3.201.
[10] Action-W User's Guide. (Version 2.0). Ardsley NY: Ambulatory monitoring. Inc.; 2001.
[11] Jean-Louis G, Kripke DF, Mason WJ, et al. Sleep estimation from wrist movement quantified by different actigraphic modalities. J Neurosci Methods 2001;105(2):185–91. https://doi.org/10.1016/S0165-0270(00)00364-2.
[12] El Baz M, Quera-Salva M, Oakley N, et al. Evaluation of actiwatch actimeter vs polysomnography in 29 patients with obstructive sleep apnea syndrome. J Sleep Res 1998;7(Suppl 2):75.
[13] Oakley NR. Validation with polysomnography of the sleep-watch sleep/wake scoring algorithm used by the actiwatch activity monitoring system. Technical report to Mini Mitter Co., Inc; 1997.
[14] Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep 2007;30(4):519–29. https://doi.org/10.1093/sleep/30.4.519.
[15] Whiting WL, Murdock KK. Emerging adults' sleep patterns and attentional capture: the pivotal role of consistency. Cognit Process 2016;17(2):155–62. https://doi.org/10.1007/s10339-016-0754-9.
[16] Boyne K, Sherry DD, Gallagher PR, et al. Accuracy of computer algorithms and the human eye in scoring actigraphy. Sleep Breath 2013;17(1):411–7. https://doi.org/10.1007/s11325-012-0709-z.
[17] Chow CM, Wong SN, Shin M, et al. Defining the rest interval associated with the main sleep period in actigraph scoring. Nat Sci Sleep 2016;8:321–8. https://doi.org/10.2147/nss.s114969.
[18] Berger AM, Wielgus KK, Young-McCaughan S, et al. Methodological challenges when using actigraphy in research. J Pain Symptom Manag 2008;36(2):191–9. https://doi.org/10.1016/j.jpainsymman.2007.10.008.
[19] Gaina A, Sekine M, Chen X, et al. Validity of child sleep diary questionnaire among junior high school children. J Epidemiol 2004;14(1):1–4.
[20] Owens JA, Spirito A, McGuinn M, et al. Sleep habits and sleep disturbance in elementary school-aged children. J Dev Behav Pediatr 2000;21(1):27–36.
[21] Maglione JE, Liu L, Neikrug AB, et al. Actigraphy for the assessment of sleep measures in Parkinson's disease. Sleep 2013;36(8):1209–17. https://doi.org/10.5665/sleep.2888.
[22] Baron KG, Duffecy J, Berendsen MA, et al. Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. Sleep Med Rev 2017. https://doi.org/10.1016/j.smrv.2017.12.002.
[23] Mantua J, Gravel N, Spencer R. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors 2016;16(5):646. https://doi.org/10.3390/s16050646.
[24] Littner M, Kushida CA, Anderson WM, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. Sleep 2003;26(3):337–41. https://doi.org/10.1093/sleep/26.3.337.
[25] Patel SR, Weng J, Rueschman M, et al. Reproducibility of a standardized actigraphy scoring algorithm for sleep in a US Hispanic/Latino population. Sleep 2015;38(9):1497–503. https://doi.org/10.5665/sleep.4998.
[26] Boudebesse C, Geoffroy PA, Bellivier F, et al. Correlations between objective and subjective sleep and circadian markers in remitted patients with bipolar disorder. Chronobiol Int 2014;31(5):698–704. https://doi.org/10.3109/07420528.2014.895742.
[27] Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002;156(9):871–81. https://doi.org/10.1093/aje/kwf113.
[28] Eidelman P, Gershon A, Kaplan K, et al. Social support and social strain in inter-episode bipolar disorder. Bipolar Disord 2012;14(6):628–40. https://doi.org/10.1111/j.1399-5618.2012.01049.x.
[29] Boudebesse C, Leboyer M, Begley A, et al. Comparison of five actigraphy scoring methods with bipolar disorder. Behav Sleep Med 2013;11(4):275–82. https://doi.org/10.1080/15402002.2012.685997.
[30] Lauderdale DS, Knutson KL, Yan LL, et al. Sleep duration: how well do self-reports reflect objective measures? The CARDIA Sleep Study. Epidemiology (Cambridge, Mass.) 2008;19(6):838–45. https://doi.org/10.1097/EDE.0b013e318187a7b0.
[31] Pesonen AK, Räikkönen K, Paavonen EJ, et al. Sleep duration and regularity are associated with behavioral problems in 8-year-old children. Int J Behav Med 2010;17(4):298–305. https://doi.org/10.1007/s12529-009-9065-1.
[32] Ustinov Y, Lichstein KL. Actigraphy reliability with normal sleepers. Behav Sleep Med 2013;11(5):313–20. https://doi.org/10.1080/15402002.2012.688779.
[33] Saksvik-Lehouillier I, Saksvik SB, Dahlberg J, et al. Mild to moderate partial sleep deprivation is associated with increased impulsivity and decreased positive affect in young adults. Sleep 2020. https://doi.org/10.1093/sleep/zsaa078.
[34] Saksvik-Lehouillier I, Langvik E, Saksvik SB, et al. Cognitive and affective functioning as a result of work onset after partial sleep deprivation. 2018. Retrieved from, https://www.researchgate.net/publication/327572903_Cognitive_and_Affective_Functioning_as_a_Result_of_Work_Onset_after_Partial_Sleep_Deprivation.
[35] Bjorvatn B. Søvndagbok. 2010. Retrieved from, https://helse-bergen.no/seksjon/sovno/Documents/sovndagbok_ny_2010.pdf.
[36] Landers R. Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. The Winnower 2015;8:e143518.81744. https://doi.org/10.15200/winn.143518.81744.