Sleep Medicine 13 (2012) 743–751

Contents lists available at SciVerse ScienceDirect

Sleep Medicine

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

Original Article

Algorithms for using an activity-based accelerometer for identification of infant sleep–wake states during nap studies

Barbara C. Galland [a,c,*], Gavin J. Kennedy [b], Edwin A. Mitchell [c], Barry J. Taylor [a]

[a] Department of Women's and Children's Health, University of Otago, Dunedin, New Zealand
[b] Department of Physical Education, University of Otago, Dunedin, New Zealand
[c] Department of Paediatrics, University of Auckland, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 4 October 2011
Received in revised form 27 January 2012
Accepted 28 January 2012
Available online 28 April 2012

Keywords:
Accelerometry
Algorithm
Arousal
Infant
Polysomnography
Sleep
Wake

ABSTRACT

Objective: To determine the accuracy of using different algorithms on the output from an Actical accelerometer, a device normally used to measure physical activity, to distinguish sleep from wake states.
Methods: Thirty-one infants aged 10–22 weeks wore the accelerometer on the shin for a daytime nap recording in tandem with polysomnography. Sleep–wake epochs were identified using four computations/algorithms: the zero-threshold computation, two common algorithms used for wrist-based devices (Sadeh and Cole), and a new algorithm developed for this study (count-scaled). Accuracy was examined in direct epoch comparison with polysomnography using 15-, 30- and 60-s sampling epochs.
Results: Overall agreements (accuracy) for sleep–wake states were >80% for all computations. The count-scaled algorithm sampling 15-s epochs gave the highest accuracy, with sensitivity (sleep agreement) at 86% and specificity (awake agreement) at 85%. Other computations yielded higher sensitivity at the expense of specificity. Another way to assess the accuracy of identification of sleep–wake states was to compare sleep parameter outputs. All computations and sampling epochs were significantly correlated with total sleep time ($r = 0.76–0.88$), sleep latency ($r = 0.70–0.93$), sleep efficiency ($r = 0.76–0.87$), and wake time after sleep onset ($r = 0.41–0.53$). The number of awakenings after sleep onset was overestimated by accelerometry.
Conclusions: The Actical accelerometer, designed to measure physical activity, can reliably identify sleep in infants during napping, with the count-scaled algorithm showing some advantages over other methods for accurate identification of sleep–wake epochs.

1. Introduction

The use of core (trunk) accelerometry is increasing in research and clinical practice with the focus on physical activity as an important health behaviour. Core accelerometers are traditionally worn on the trunk, and activity data are used to work out energy expenditure. It would be very useful if the same recordings from activity-based devices could also be used to estimate sleep. Activity, or, rather, inactivity, data from an accelerometer worn on the wrist of the non-dominant hand are traditionally used to estimate sleep; this is often called "actigraphy" by users of actigraph devices. The "gold-standard" method for measuring sleep is polysomnography (PSG), but this has the disadvantages of being resource intensive, intrusive, and (usually) laboratory based.

Activity-based core accelerometers are potential tools to advance research investigations and clinical practice concerning the link between physical exercise, sleep, and health. Examples of research investigations include studies showing that physical exercise contributes to healthy sleep patterns across different age groups [1,2], and the growing body of epidemiological studies showing a link between short sleep duration and the development of obesity [3–6]. Regarding the latter, detailed studies using accelerometry have shown associations between later sleep onset, shorter sleep time, more disrupted sleep (frequent night waking), daytime sleepiness [7], irregular sleep, and short sleep duration and variability with being overweight or obese [8]. Research has advanced to clinical trials with a focus on weight management through sleep intervention [9–11].

The sleep measurement capabilities of activity-based core accelerometers have received little attention. Core accelerometry devices and wrist devices differ in several ways. For example, the core Actical accelerometer (Mini-Mitter Co., Inc., Bend, OR, USA) used in the current study and its wrist counterpart (Actiwatch, Mini-Mitter Co., Inc.) use similar technology; both devices use

* Corresponding author at: Department of Women's and Children's Health, University of Otago, P.O. Box 913, Dunedin, New Zealand. Tel.: +64 3 474 0999; fax: +64 3 474 7817.
E-mail address: barbara.galland@otago.ac.nz (B.C. Galland).

1389-9457/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sleep.2012.01.018

motion sensors but, oriented differently inside the devices. In the Actical accelerometer, the sensor is orientated to detect more vertical acceleration associated with walking/running, whereas the sensor in the Actiwatch is more sensitive to smaller movements of the lower arm. The Actiwatch is set to a higher sensitivity to detect small movements indicative of wakefulness [12]. The Actical accelerometer provides an objective quantifiable measure of sedentary and physical activity and energy expenditure. Weiss et al. [13] recently showed that an Actical accelerometer worn on the wrist can reliably measure sleep duration, but not sleep efficiency, in adolescents; further validation studies are required to investigate its use for sleep estimation addressing several factors such as placement, suitability for use across different age groups, and clinical vs normal populations.

To study sleep and wake patterns in infants, 24-h recordings are desirable. The Actical accelerometer is small, water resistant, and stores information for up to 45 days using 1-min selected epochs. To the authors' knowledge, only one other study has employed core accelerometry for the identification of sleep–wake states in infants [14]. An unspecified accelerometer attached to the infant's nappy was claimed to be a reliable method for estimating sleep–wake states, with approximately 77–92% accuracy compared with PSG. Several other studies have used wrist devices (worn on the leg or ankle) to estimate sleep in infants, and this method has been validated against PSG and behavioural sleep–wake state scoring [15–17].

The purpose of this study was to compare activity-based Actical accelerometer outputs with "gold standard" PSG during infant nap fragmentation studies using a protocol mimicking sleep disturbance, as one deficiency of many wrist devices has been accurate identification of waking after sleep onset (WASO). More specifically, this study aimed to identify the best performing computation/algorithm, and sampling epoch, for the identification of sleep–wake states from those commonly used with wrist devices (zero-threshold computation [18], Sadeh [16] and Cole [19] algorithms) and from a new algorithm developed for this study (count-scaled algorithm). Finally, this study sought to identify the best performing computation/algorithm with regards to measuring common sleep parameters.

2. Methods

This study was part of a larger study investigating the arousability of three-month-old infants. The last 33 participants of the larger study wore the Actical accelerometer and combined it with PSG. Of these 33 subjects, 24 were male and nine were female, median age was 13.0 weeks (range 10.0–22.3 weeks), and average weight was 6.33 kg (range 5.0–8.1 kg). Eighteen infants were formula fed, 10 were exclusively breast fed, and five were fed both breast milk and formula. Sixteen infants used a dummy to go to sleep and 17 did not. All infants were placed on their back to sleep. Informed written consent was obtained from the infant's parent/s or caregiver/s. The Northern Regional Ethics Committee approved the study protocol (NT/06/06/068).

2.1. Procedures

Infants were brought into the sleep laboratory to set-up approximately 1 h before their normal nap time. A multichannel polysomnograph (Compumedics E series, Australia) was used to record the following: electro-encephalogram (EEG; C4/A1, C3/A2), electro-oculogram (ROC, LOC); electromyogram (submental); electrocardiogram (modified lead II position); respiratory pattern (thoracic and abdominal respibands) and pulse oximetry (Masimo Radical 7 monitor; Masimo Corp, Irvine, CA, USA). Simultaneous audio/video monitoring was digitally recorded. The Actical accelerometer was secured around the infant's shin at the midpoint between the knee and ankle with the arrow pointing towards the head. Immediately prior to placement, the Actical accelerometer was synchronized to the polygraph trace time.

2.2. Auditory stimulation

The infant was placed supine in a cot with the top of the head approximately 6 cm from a speaker that delivered white noise over a range over 50–100 dB. Once sleep was firmly established, the noise was administered, beginning at 50 dB and incrementally increased by 10 dB every 90 s until cortical arousal or full awakening, or until 100 dB was reached. Test sequences were repeated to obtain two threshold levels for each sleep state. The duration of the stimulus was 3 s. At the end of auditory testing, infants were left to wake naturally.

2.3. Polysomnography

2.3.1. Polysomnographic sleep–wake criteria

Based on the PSG recordings, sleep stages were scored as quiet sleep (QS), active sleep (AS), and indeterminate sleep (IS) in 30-s epochs according to the criteria adopted by Anders et al. [20]. For analysis purposes, all QS, AS, and IS epochs were coded as sleep. Thirty-second epochs were divided into 15-s epochs for direct comparison with the accelerometer output, and dichotomized into categories 0 (sleep) and 1 (wake). Full awakenings were marked as an abrupt shift in EEG frequency according to the Atlas Task Force criteria [21] and meeting the criteria for consensus scoring of awakenings in infants [22]. Awakenings followed a cortical arousal and lasted longer than 15 s. These were distinguished from a cortical arousal, which is the occurrence of an abrupt change in EEG background frequency of at least 1 Hz for a minimum of 3 s with at least two of the following changes: a gross body movement detected by movement sensors or seen as artefact movement in the somatic channels, changes in heart rate (at least 10% of baseline values), or changes in breathing pattern in terms of frequency or amplitude [22].

2.4. Accelerometry

The Actical accelerometer has dimensions of $8 \times 27 \times 10$ mm and weighs 17.5 g. It is sensitive to movements in the range of 0.5–3 Hz, samples at 32 Hz and has a memory capacity of 64 kB. Actical data were recorded at the highest resolution (15-s epochs), and the Actical energy expenditure data were exported from the software program in spreadsheet format for importation into MATLAB (MathWorks Inc., MA, USA).

2.5. Algorithms for accelerometer sleep–wake identification

The zero-threshold computation [18] and three algorithms (Sadeh [16], Cole [19], and count-scaled algorithms) were tested for the identification of sleep–wake states from the Actical data All algorithms were written in MATLAB (MathWorks Inc.) and outputs were computed using this software system.

2.5.1. Zero-threshold computation [18]

This method derives the amount of time per epoch that the activity is above the threshold value (set at zero), such that epochs >0 are scored as wake and epochs = 0 are scored as sleep.

2.5.2. Sadeh algorithm [16]

The Sadeh algorithm is computed as follows:

$$PS = 7.601 - 0.065MW5 - 1.08NAT - 0.056SD6 - 0.073\ln(ACT)$$

where PS is the probability of sleep; MW5 is the average number of activity counts during the scored epoch and a window of five epochs preceding and following the scored epoch; NAT is the number of epochs with an activity level of $\geqslant 50$ but <100 activity counts in an 11-min window, including the scored epoch and the five epochs preceding and following the scored epoch; SD6 is the standard deviation of the activity counts during the scored epoch and the five preceding epochs; and ln(ACT) is the natural logarithm of the number of activity counts during the scored epoch + 1. If PS $\geqslant 0$, the epoch is scored as sleep; otherwise, it is scored as wake.

2.5.3. Cole algorithm [19]

The Cole algorithm computes a weighted sum of the activity in the current minute, the preceding 4 min, and the following 2 min as follows:

$$S = 0.0033(1.06an4 + 0.54an3 + 0.58an2 + 0.76an1 + 2.3a0 + 0.74al + 0.67a2)$$

where $an4$–$an1$ are activity counts from the prior 4 min, $a0$ is the current minute, and $a1$ and $a2$ are the following 2 min. The current minute is scored as sleep when $S < 1$.

2.5.4. Count-scaled algorithm

The count-scaled algorithm, devised by a co-author (GK), is very similar to the Cole algorithm [19] excepting that it is performed using count-scaled data. Each trial (entire recording period of each participant) is scaled relative to the mean value of all epochs that have non-zero counts. This makes the algorithm able to cope with different accelerometers (with different count thresholds or placements, etc.) that may reduce/increase the number of counts per epoch. Sole use of the non-zero counts prevents sustained periods of zero activity from reducing the apparent sensitivity of the accelerometer. In Fig. 1, the mean activity count in non-zero epochs for the whole trial was determined to be 30. All epoch counts are therefore divided by 30. Once this stage has been completed, the algorithm operates in a similar way to the Cole algorithm [19].

The optimization (described below) of the count-scaled algorithm parameters started from the weighting factor (W; described below) values, $W(1–7)$, of [1 2 3 4 5 3 1] and scaling factor (S) of 5. These starting parameters[1] were chosen through a small series of "trial and error" calculations using a few randomly selected trials of data. The optimization process for the raw 15-s data reached values of: $W(1) = 1.17$, $W(2) = 1.09$, $W(3) = 2.57$, $W(4) = 4.30$, $W(5) = 5.05$, $W(6) = 4.01$, $W(7) = 0.82$ and $S = 2.7$. There was very little performance improvement (change of 0.2% agreement) from the case when the scaling factor, S, alone was optimized; i.e., when $W(1–7)$ was fixed at the values [1 2 3 4 5 3 1].

Optimization refers to adjusting the values of the constants that make up an algorithm to minimize the sum square difference between its output and the desired output. The difference used in the optimization process is the sum for all trials. When adjustments are made to the values of the constants, any resulting change in the difference to PSG will affect the magnitude and direction (positive or negative) of any proceeding changes to the constants. The process is repeated as long as is necessary until no more optimization occurs. Due to the binary nature of the PSG and sleep–wake data, there are often large ranges for the values of the constants yielding the same differences. This caused a number of problems when using standard optimization techniques, as false minima/plateaus were frequently reached.

Weighting factors determine the "shape" of the data filter. These are used to scale each of the seven epochs in the algorithm and offer a way to indicate the relative importance of each epoch in determining the sleep–wake state of the current epoch. For example, in Fig. 1, the current epoch is a strong indicator and therefore has a weight of 5, the previous epoch has a slightly weaker indicator and therefore has a weight of 4, the epoch before that has a weight of 3, etc.

The scaling factor scales the sum of all the weighted epochs to the value, which is then compared with the sleep–wake threshold of 1.

2.6. Sleep–wake identification using different sampling epochs

Many algorithms for the identification of sleep–wake states resample the raw accelerometry (activity count) data into 60-s epochs. The accuracy of the raw 15-s epochs and those resampled into 30- and 60-s epochs were tested using moving averages. For 30-s epochs, the average of two (current and next) adjoining epochs was used to replace each value. For 60-s epochs, the average of four (previous, current, and next two) adjoining epochs was used to replace each value. Using these methods, the total number of epochs to compare accelerometry vs PSG outputs remained the same as for the raw 15-s counts.

2.7. Sleep parameters

Standard sleep parameters were calculated using the zero-threshold criteria and the three algorithms in comparison with PSG: sleep latency, total sleep time (TST), sleep efficiency, wake time after sleep onset (WASO [duration]) and number of awakenings after sleep onset (WASO [number]). TST was calculated as the total duration of epochs scored as sleep between lights out and lights on. Sleep efficiency was calculated as the ratio of TST to the total time between lights out and lights on. Sleep latency was the time from lights out to sleep onset. An awakening during PSG and activity algorithm is defined as when the sleep–wake state is scored as awake for $\geqslant 30$ consecutive seconds after sleep onset and before the end of data recording.

2.8. Data analysis

When comparing the methods (accelerometer vs PSG), PSG was considered to be the gold standard. Standard methods were used to compute the overall agreement (%), sensitivity (% sleep agreement), and specificity (% wake agreement). Agreement accuracy was determined by simple kappa and by prevalence- and bias-adjusted kappa (PABAK). The level of agreement was interpreted based on the Landis and Koch scale [23]: $\leqslant 0$, poor agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement. As this was a nap study, epochs comprising the total recording time were highly skewed towards more sleep epochs. To provide equal weights to sleep and waking episodes, PABAK [23] was computed to counteract bias. For sleep parameters, Pearson's correlation coefficients were calculated following log-transformation of skewed variables. The practice of Insana et al. [25] was adopted to produce PSG and Actical output (count-scaled algorithm) plots for the five sleep parameters. The plots are similar to Bland–Altman [26] agreement plots

[1] The starting parameters were calculated for three algorithm structures, using three-, five-, seven-, and nine-epoch weightings. MATLAB (MathWorks Inc.) was used to try every combination of weights (in whole-number increments from 1 to 5) and scaling factor (in whole-number increments from 1 to 10). This simplistic "trial and error" approach took a considerable amount of computation time (with over 19 million value combinations for the nine-epoch case), and so was only performed on a small number of randomly selected trials. The best performing structure, seven epochs with weights $W[1–7] = [1\,2\,3\,4\,5\,3\,1]$ and scaling factor $S = 3$, was used as the basis of the subsequent optimization on all of the trials.

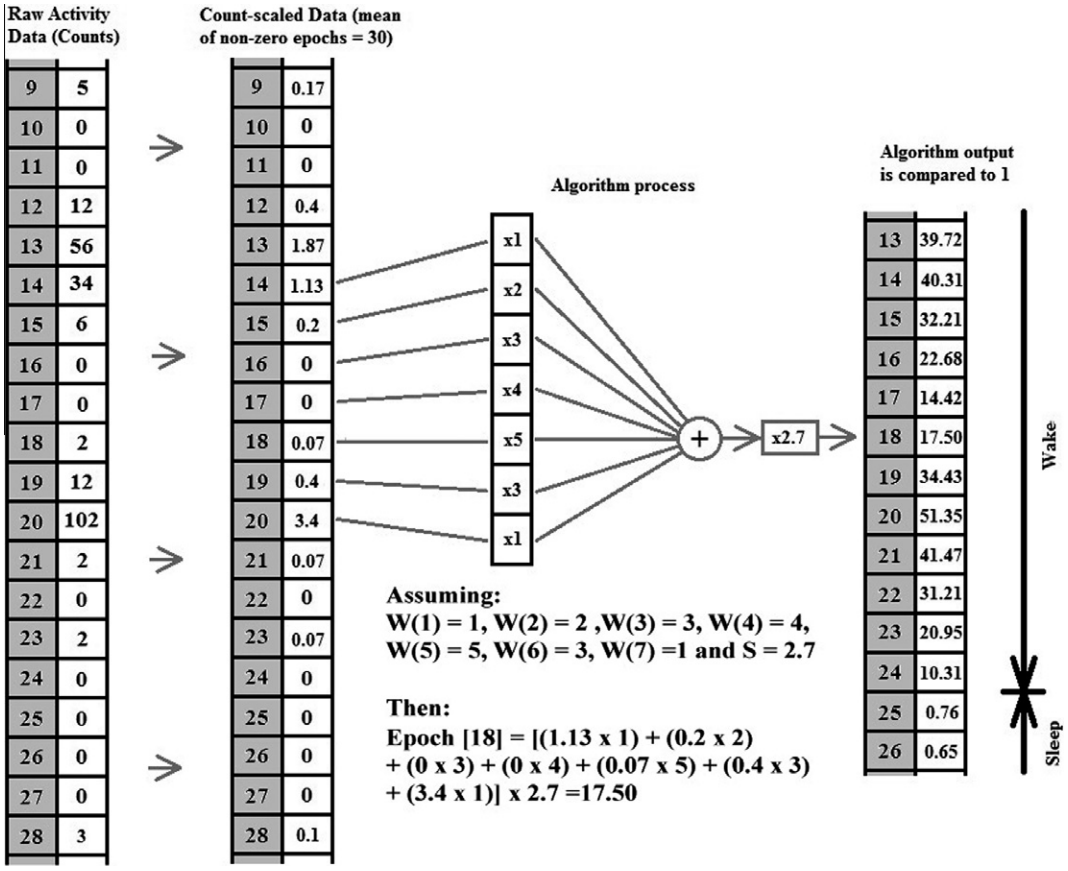

Fig. 1. Diagram of the mathematical process leading to sleep–wake outputs using the count-scaled algorithm. Shaded blocks contain consecutive epoch numbers. *W*, weighting factors; *S*, scaling factor. Note: algorithm output ⩾1 = wake and <1 = sleep. See text for explanation of how *W* and *S* were calculated.

but only the average PSG value is plotted on the *x*-axis, whereas the Bland–Altman method plots the average of the PSG and Actical values. As the Actical is not assumed to be a valid measure in the strict context of the Bland–Altman method, the *x*-axis without Actical data excludes the Actical error in plot interpretation. The *y*-axis is the difference between concurrent PSG and Actical measures. *P*-values <0.05 were considered to be statistically significant. All statistical analyses were conducted using Stata Version 11 (StataCorp, College Station, TX, USA).

3. Results

The recordings for two infants were discarded from the analysis because of faulty accelerometer outputs; both recordings were from the same accelerometer. Thus, data from 31 infants were used in the final analysis. Sleep data were representative of both AS and QS in all but one infant (QS only).

3.1. Sleep–wake agreement

The overall agreement, sensitivity, specificity, and kappa statistics of the epoch-by-epoch comparison produced from raw (15-s epochs) and resampled data (30- and 60-s epochs) are reported in Table 1. Overall agreement of >80% was reached with all computations/algorithms tested under all conditions. Sensitivity for sleep agreement was >90% for three methods sampled at 15- and 30-s epochs (zero-threshold computation, Sadeh and Cole algorithms). However, this was at the expense of wake agreement, with specificity ranging from 57% to 80%. The 60-s outputs switched the balance; wake agreement became higher

Table 1
Accuracy of Actical accelerometry compared with polysomnography for the identification of sleep–wake states using the four computations/algorithms at different sampling epochs.

Sampling epoch	Computation/algorithm	Overall agreement	Sensitivity	Specificity	Kappa	PABAK
15 s	Zero-threshold	82.3 (78.7–85.8)	94.8 (90.6–99.1)	57.8 (50.4–65.2)	0.55 (0.48–0.62)	0.65 (0.57–0.72)
	Sadeh	84.3 (81.1–87.6)	95.1 (92.0–98.3)	63.1 (55.5–65.2)	0.59 (0.53–0.66)	0.69 (0.62–0.75)
	Cole	83.2 (79.9–86.6)	94.6 (91.1–98.8)	60.2 (52.8–67.6)	0.57 (0.50–0.64)	0.66 (0.59–0.73)
	Count-scaled	86.3 (84.1–88.4)	85.7 (82.4–89.1)	84.3 (79.5–88.7)	0.66 (0.59–0.72)	0.72 (0.68–0.77)
30 s	Zero-threshold	86.3 (83.9–88.8)	91.0 (88.4–93.5)	77.7 (71.8–83.8)	0.65 (0.59–0.71)	0.72 (0.68–0.78)
	Sadeh	86.8 (84.6–89.0)	90.8 (88.3–93.4)	79.0 (73.2–84.9)	0.66 (0.61–0.72)	0.74 (0.70–0.78)
	Cole	86.5 (84.2–88.9)	90.9 (88.3–93.4)	78.2 (72.4–84.2)	0.66 (0.60–0.72)	0.73 (0.68–0.77)
	Count-scaled	86.1 (84.1–88.3)	84.7 (81.4–88.0)	87.4 (83.2–91.8)	0.66 (0.60–0.72)	0.72 (0.68–0.77)
60 s	Zero-threshold	84.5 (82.1–86.9)	80.0 (75.8–84.3)	89.9 (85.9–93.8)	0.63 (0.57–0.69)	0.69 (0.64–0.74)
	Sadeh	84.6 (82.2–87.0)	79.9 (75.7–84.2)	90.3 (86.4–94.2)	0.63 (0.57–0.69)	0.69 (0.64–0.74)
	Cole	84.5 (82.1–86.9)	80.0 (75.7–84.2)	89.9 (86.0–93.9)	0.63(0.57–0.69)	0.69 (0.64–0.74)
	Count-scaled	84.5 (82.1–86.9)	79.8 (75.7–84.0)	90.2 (86.4–94.0)	0.63 (0.57–0.69)	0.69 (0.64–0.74)

Sensitivity, sleep agreement; specificity, wake agreement; PABAK, prevalence- and bias-adjusted kappa.

Table 2
Sleep parameters calculated from the different computations/algorithms using data sampled at 15-s epochs.

Computation/algorithm	Sleep latency (min)				Wake after sleep onset (number)				Wake after sleep onset (duration, min)				Total sleep time (min)				Sleep efficiency (%)			
	Mean (SD)	Median (IQR)	r^a	P	Mean (SD)	Median (IQR)	r^a	P	Mean (SD)	Median (IQR)	r^a	P	Mean (SD)	Median (IQR)	r^a	P	Mean (SD)	Median (IQR)	r^a	P
Polysomnography	17 (14)	15 (7–30)			1 (0)	1 (0–2)			3.5 (5.5)	2 (0–3)			39 (13)	35 (31–48)			63.7 (19.4)	68.9 (47.6–77.1)		
Zero-threshold	12** (12)	7 (2–19)	0.80	<0.001	4** (3)	4 (3–5)	0.38	0.035	10.4** (9.4)	7 (4–14)	0.48	0.006	50* (15)	41 (38–58)	0.76	<0.001	80.2** (13.5)	82.8 (72.3–89.4)	0.76	<0.001
Sadeh	17 (14)	17 (5–32)	0.93	<0.001	3** (2)	3 (2–5)	0.13	0.475	6.9** (5.9)	5 (3–11)	0.50	0.004	48* (15)	41 (36–55)	0.80	<0.001	77.3* (14.9)	78.0 (68.0–9.4)	0.81	<0.001
Cole	13* (13)	8 (2–24)	0.78	<0.001	4** (2)	4 (3–5)	0.29	0.110	10.4** (11.0)	6 (4–15)	0.41	0.021	49* (15)	41 (38–57)	0.78	<0.001	78.9* (14.0)	80.1 (70.3–89.4)	0.79	<0.001
Count-scaled	14* (13)	9 (2–25)	0.79	<0.001	4** (2)	3 (3–5)	0.35	0.057	10.0** (10.8)	7 (4–11)	0.48	0.006	38 (13)	34 (28–45)	0.83	<0.001	59.8 (16.6)	57.4 (50.9–73.3)	0.87	<0.001

SD, standard deviation; IQR, interquartile range.
Values represent median and IQR.
[a] Pearson correlation coefficient with polysomnography.
* $P < 0.05$, paired t-test, algorithm parameter outcome significantly different from that of polysomnography.
** $P < 0.001$, paired t-test, algorithm parameter outcome significantly different from that of polysomnography.

Table 3
Pearson correlation coefficients for sleep parameters comparing polysomnography with Actical outputs.

Sampling epoch (s)	Sleep latency			Wake after sleep onset (number)			Wake after sleep onset (duration)			Total sleep time			Sleep efficiency		
	15	30	60	15	30	60	15	30	60	15	30	60	15	30	60
Zero-threshold	0.80***	0.70***	0.88***	0.38*	0.19	0.15	0.48**	0.44*	0.53**	0.76***	0.87***	0.87***	0.76***	0.87***	0.85***
Sadeh	0.93***	0.70***	0.88***	0.13	0.14	0.15	0.50**	0.45*	0.53**	0.80***	0.88***	0.87**	0.81***	0.88***	0.85***
Cole	0.78***	0.70***	0.88***	0.29	0.18	0.15	0.41*	0.45*	0.53**	0.78***	0.88***	0.87***	0.79***	0.88***	0.85***
Count-scaled	0.79***	0.76***	0.88***	0.35	0.16	0.16	0.48**	0.50**	0.52**	0.83***	0.85***	0.87***	0.87***	0.86***	0.85***

* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

at the expense of sleep agreement. The best-performing algorithm, giving the highest possible sensitivity (86%) with the highest possible specificity (84%), was the count-scaled algorithm at 15-s epochs; the kappa statistic classed agreement as "substantial" [23]. This improved with bias removal (PABAK = 0.72).

3.2. Sleep parameters: correlation and agreement

Sleep parameters were tested using all computations/algorithms. Sleep parameter data outputs using the 15-s raw sampling epoch are given in Table 2. All four methods produced sleep latency, TST, and sleep efficiency measures for accelerometry that were significantly correlated with PSG, but sleep latency was underestimated (excluding the Sadeh algorithm that produced similar values to PSG), and TST and sleep efficiency were overestimated (excluding the count-scaled algorithm that produced similar values to PSG). For WASO (number), the zero-threshold computation was significantly correlated with PSG, the count-scaled algorithm showed this tendency, but the Cole and Sadeh algorithms were not correlated. WASO (number) was overestimated for all. For WASO (duration), outputs from

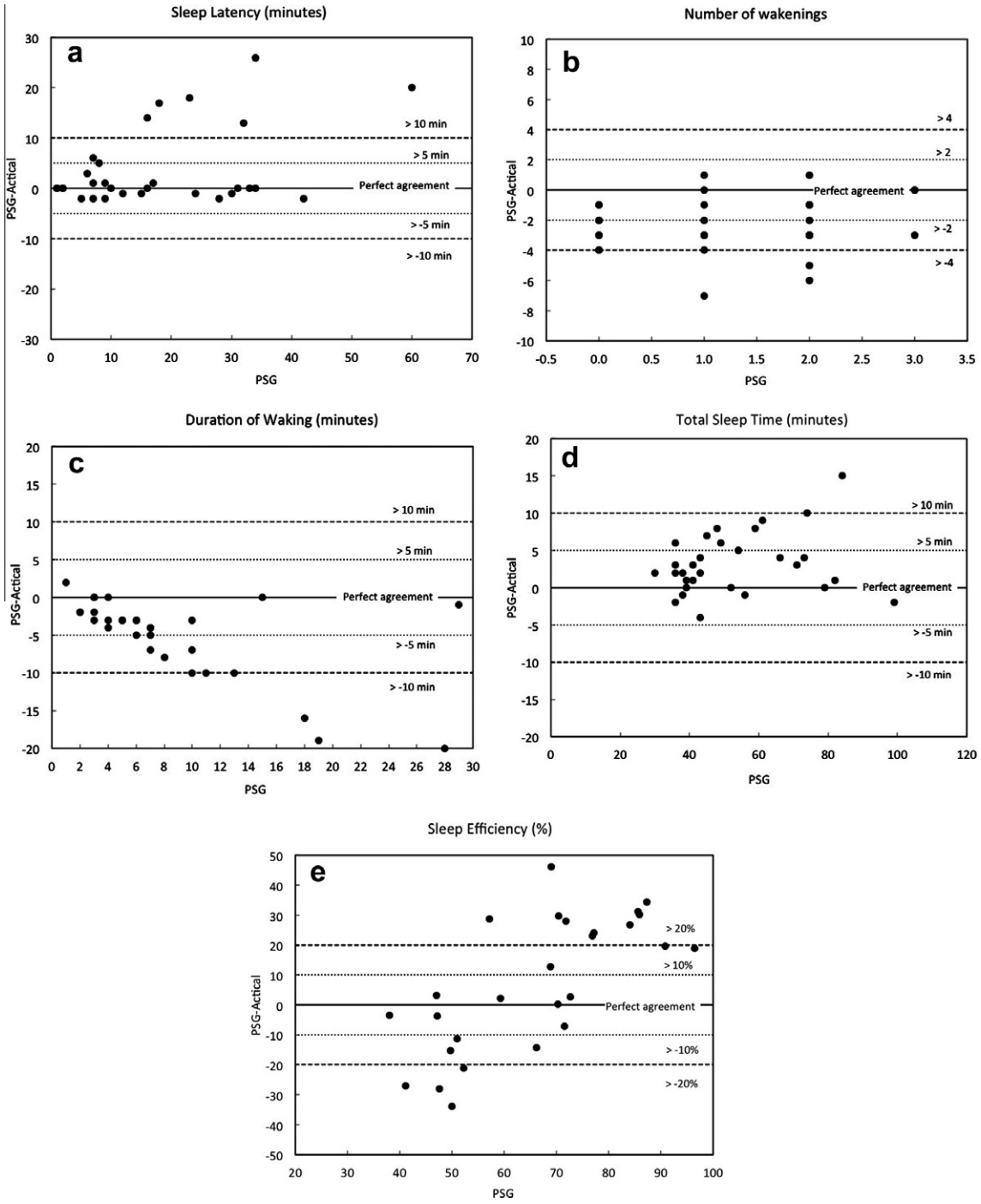

Fig. 2. Bland–Altman plots of polysomnography (PSG) and accelerometry sleep measures using the Sadeh algorithm. The dotted lines reveal the mean and upper and lower agreement limits.

all computations/algorithms were correlated with PSG, but duration was significantly overestimated.

Table S1 (see online Supplementary material) provides significance levels for the differences between all computations/algorithms and PSG for the five sleep parameter outputs. Highly significant differences were found for both WASO measures for all computations/algorithms compared with PSG. The close comparison of the count-threshold algorithm with the Cole algorithm is evident, with four sleep parameters producing similar outputs, excluding sleep efficiency.

Table 3 summarizes the correlations for all sampling epochs. For all parameters, as the sampling epoch increased, there was negligible difference in computation/algorithm performance. Increasing the duration of the sampling epoch resulted in poorer correlations for WASO (number) using 60-s epochs.

Fig. 2a–e plots the five variables in the correlation analysis for the count-scaled algorithm (15-s epoch) to show the agreement and spread of bias over the value ranges. In the majority of infants (n = 24, 77%), sleep latency (Fig. 2a) calculated by the count-scaled algorithm was within ±5 min of the PSG calculation, and the results were in perfect agreement for 10 infants (32%). In seven infants, sleep latency was underestimated by more than 5 min. For WASO (number), the calculations for three infants were in perfect agreement, the count-scaled algorithm overestimated WASO in 27 infants (87%), and the count-scaled algorithm underestimated WASO in two infants (Fig. 2b). WASO (duration) was also overestimated in the majority of infants (n = 26, 84%), with just four infants having values in perfect agreement (Fig. 2c). A downward displacement is obvious in the WASO (duration), tending to show more error (overestimation) with longer awakenings. This observation may link with the algorithm having poorer specificity than sensitivity. TST was underestimated in 23 infants (75%) and overestimated in five infants (16%); values were in perfect agreement for three infants. Sleep efficiency was highly variable in terms of agreement, showing wide spread in the ability of the count-scaled algorithm to both over- and underestimate sleep (±>20%) in comparison with PSG (Fig. 2e).

4. Discussion

The results of this study support the use of activity-based accelerometry in the identification of sleep–wake states in infants during fragmented napping. The common sleep computations/algorithms (zero-threshold computation, Sadeh and Cole algorithms) showed "substantial" agreement [23] using raw 15-s epochs, or 30- or 60-s moving average epochs. The newly developed count-scaled algorithm gave the highest level of agreement in terms of both sleep and wake agreement using 15-s epochs, and it is suggested that future studies using the Actical accelerometer to estimate infant sleep–wake states could use this algorithm. However, as these agreements were achieved using an infant nap fragmentation study protocol, it remains to be determined whether or not the accuracy holds for overnight sleep. Only a few other devices have been validated for infant sleep against PSG: the AMA-32 (Ambulatory Monitoring Inc.) [16], the Actiwatch AW64 (Mini Mitter Co., Inc.) [15,27], and an unnamed accelerometer [14]. The Actical accelerometer used in the present study has the added advantage of combining 24-h energy expenditure data in the recording.

The main benefit of introducing the count-scaled algorithm is that it is independent of the number of counts per epoch. Using the mean counts instead of the maximum counts prevents a few anomalously large data spikes from affecting the shape of the whole trial. Provided the relative "shape" of the activity graph is similar between accelerometers and placement location, the output variables should be the same, although this needs to be confirmed by further testing. The Sadeh sleep-scoring algorithm has been validated for 1-year-old children wearing the accelerometer on the ankle, and for older children and adolescents wearing the accelerometer on the non-dominant wrist [16,28]. Sadeh et al. [16] suggested that different algorithms may need to be used for different ages. The present data were restricted to infants aged approximately three months, and these findings may not necessarily apply to other age groups.

All computations/algorithms yielded strong correlations between PSG and accelerometry for sleep latency, TST, and sleep efficiency, with weaker correlations for WASO (duration). WASO (number) was poorly correlated, with the accelerometer overestimating this parameter. Accelerometer sensitivity may not be sufficiently high to detect small movements that occur while the infant is lying quietly awake. However, the zero-threshold computation and the count-scaled algorithm produced the best correlations at 15-s epochs (0.38 and 0.35, respectively), suggesting greater accuracy of this computation/algorithm when raw data counts are used.

The wrist counterpart of the Actical accelerometer (Actiwatch) is set to a higher sensitivity to detect small movements indicative of wakefulness. Using the Actiwatch for overnight PSG in 14-month-old infants, Insana et al. [25] reported an overall agreement of 90%, with high sleep agreement within each sleep stage. However, their wake agreements were not as strong (59%). The short-period accelerometry data in the present study also produced better sleep than wake agreement, but the outcome was much improved using the count-scaled algorithm sampling at 15-s epochs. The use of longer epochs (60 s) gave much better wake agreement, but at the cost of sleep agreement, and thus was not a viable option with which to progress. So et al. [15] reported higher overall sleep agreement using a different brand of wrist accelerometer; they reported 94% and 92% in infants aged 2–4 weeks and 5–6 months, respectively. However, their postfiltering of all transient WASO did not follow by correcting for the different duration of sleep–wake states in calculating agreement. If agreements are high or low (outside of 20% or 80%), adjustments need to be made [24]. PABAK was developed to adjust the measures of agreement for the potential prevalence and bias effect. Overall agreement of 86% was noted in the present study, with the strongest predictor of sleep and wake combined being the count-scaled algorithm. The kappa statistic allowed the authors to categorize the agreement beyond chance to report that agreements between accelerometry and PSG were "substantial" and improved with bias removal.

Few studies have used core accelerometer devices with energy expenditure capabilities to additionally identify sleep–wake states. Weiss et al. [13] used the 20-count threshold criteria (time above threshold) to estimate sleep in 30 adolescents, comparing Actical accelerometry with overnight PSG, and reported a weak correlation with sleep efficiency (0.348) but a better correlation with TST (0.722), closer to the range reported in the present study. The 20-count threshold was tried unsuccessfully in the present study, probably because the threshold level is only valid for certain accelerometers and body locations. When the threshold level was optimized for best sleep–wake classification, it tended towards zero. In the case of optimizing a single threshold value, when the optimization began at a non-zero value, it always tended towards zero. When the optimization was fixed so the threshold began at zero, it did not change, suggesting the zero threshold to be most suitable.

In this study, TST was highly correlated with PSG; all four methods produced correlations ranging from 0.76 to 0.83. Sleep efficiency was also very strongly correlated with PSG, ranging from 0.76 to 0.87. Wrist accelerometry devices were not included for comparison. Weiss et al. [13] reported a strong correlation be-

tween TST recorded by the activity-based Actical accelerometer worn on the wrist and two wrist devices designed to estimate sleep (Actiwatch and Sleepwatch), but a weak correlation for sleep efficiency. Unfortunately, they did not report on epoch-by-epoch sleep–wake agreement.

The zero-threshold computation and Sadeh and Cole algorithms overestimated TST. This is typical of wrist devices used in recording sleep in healthy subjects, and has also been reported when using the Actical accelerometer [13]. However, the count-scaled algorithm measured TST more accurately. Overestimations were most likely due to the infant lying quietly awake, particularly close to sleep onset, and some overestimation was reflected in the shorter sleep latency. Sadeh and Acebo [29] reported that most discrepancies between wrist accelerometry devices and PSG are during transitions to and from sleep. The present data concur with this with regards to the Actical accelerometer worn on the shin using common sleep–wake algorithms, but the count-scaled algorithm appears to show some benefit over the other algorithms, at least for this nap fragmentation study. Using a count-threshold algorithm, Insana et al. [25] reported that actigraphy becomes increasingly worse in the identification of TST as arousal frequency increases. Underestimation of TST is more likely to occur in subjects with sleep disorders particularly linked to movement disorders (e.g., periodic limb movements in children [30] or even sleep-disordered breathing in children [31]). Unlike PSG, no published guidelines for accelerometry exist for the identification of sleep–wake states for any age group. Only more standardized practices would make this possible. Currently, accelerometers vary greatly in terms of both hardware (e.g., sensitivity and specifications of the accelerometer) and software (e.g., definitions of sleep measures), and few brands include age-specific scoring criteria.

Measuring WASO is important for measuring sleep quality, but the present results indicate that further algorithm refinement is required for this measure. Although WASO (duration) was significantly correlated with PSG, the data show that the accelerometry values were consistently overestimated. It remains unknown whether or not the different outputs produced by the algorithms in relation to WASO measures are clinically significant, and this would require overnight recordings and knowledge of clinically/practically important outcomes allied to inaccuracy of WASO (e.g., daytime impairment). The illustrated plots show that the accuracy of agreement became worse as WASO (number) or WASO (duration) increased.

The main study limitation was that recordings were confined to one nap fragmentation study, and the data were obtained under experimental conditions where waking was provoked as part of an arousal study rather than conducted over a period of undisturbed sleep. Furthermore, the aims of this study were not those of the larger main study, and the protocol was not designed to identify naps within a 24-h period. With these limitations in mind, it is suggested that further studies are required to investigate wider utility of the Actical accelerometer and the count-scaled algorithm, testing accuracy over longer periods of undisturbed naps and overnight sleep, across different age groups, and between normal and clinic populations. However, it could be argued that the count-scaled algorithm may perform better when sleep is undisturbed given that: (1) the sleep agreement was high (i.e., fewer sleep–wake transitions and arousals would have to be contended with within an undisturbed sleep protocol) and (2) the accelerometer performance was comparable with other validation studies during non-interrupted sleep protocols in respect to poor specificity (wake agreement). Accelerometry wake agreement remains an issue. However, accuracy of wrist devices with regard to waking parameters (after sleep onset) can be improved when the devices are used in conjunction with a sleep log [32], and this may also apply to core accelerometer devices such as the Actical accelerometer. A recent study in children suggested that it may be possible to include one or two nights of actigraphy without a sleep log, and that event markers for sleep onset and offset can also be used when no log is present [33].

This study only tested three algorithms, but other algorithms commonly used within the sleep analysis software of various actigraphs are similar to the Cole algorithm with fixed weighting/scaling factors, and thus may have produced very similar results to the Cole algorithmm, e.g., the Cambridge algorithm for Sleepwatch (Neurotechnology Ltd.), the UCSD algorithm [34] for the Mini Motionlogger (Ambulatory Monitoring Inc.), and the threshold-based algorithm used for the Actiwatch-L (Mini-Mitter Co., Inc.). With the exception of the Sadeh algorithm and that derived from Lotjonen et al. [35], most actigraphy algorithms use a combination of a scaling factor and epoch weightings to produce an output that is compared against a fixed threshold. These algorithms act as basic filters to reduce signal/data "noise" and give a more accurate indication of sleep–wake states. These algorithms have been validated against PSG for a single (or limited) device/placement combination, and therefore using the algorithms in an "as is" approach with different device/placement arrangements could produce inaccurate results. The uniqueness of the count-scaled algorithm is in the pre-scaling of the accelerometer count data to produce an algorithm input unaffected by the magnitude of the count signal.

In conclusion, this study showed that a device designed to measure physical activity can reliably estimate sleep in infants over a single daytime nap fragmentation study, and that the new count-scaled algorithm shows some advantage over existing algorithms with respect to balancing sleep and wake agreement. However, it is not known if sleep–wake accuracy can be maintained for longer overnight sleep or 24-h recordings. Further development or adjustment is needed for these devices to estimate WASO accurately. The unique scaling aspect of the count-scaling algorithm means that it could be applied to other accelerometers where count outputs differ, either from different sensor sensitivities or from different placement on the body aligned to movement that can reduce or increase the number of counts per epoch (e.g., leg vs waist). The ability to use this single device for both activity and sleep has many advantages for epidemiological and clinical research, and the algorithm could be used for Actical archival dataset analysis. These findings represent the first step in identifying the utility of the activity-based accelerometer for infant sleep studies, with implications for wider application of the algorithm for use with other devices.

Funding source

The larger study was funded by Cure Kids in collaboration with the New Zealand IDEAL Study.

Conflict of Interest

E.A. Mitchell is supported, in part, by Cure Kids.

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2012.01.018.

Acknowledgements

The authors wish to thank the Infant Development, Environment, and Lifestyle (IDEAL) Study team for their assistance, and they would like to thank Andrew Gray for his advice on statistical analyses. The authors also wish to thank the infants and their parents/guardians who participated in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sleep.2012.01.018.

References

[1] Dworak M, Wiater A, Alfer D, Stephan E, Hollmann W, Struder HK. Increased slow wave sleep and reduced stage 2 sleep in children depending on exercise intensity. Sleep Med 2008;9:266–72.
[2] Lira FS, Pimentel GD, Santos RV, Oyama LM, Damaso AR, Oller do Nascimento CM, et al. Exercise training improves sleep pattern and metabolic profile in elderly people in a time-dependent manner. Lipids Health Dis 2011;10:1–6.
[3] Landhuis CE, Poulton R, Welch D, Hancox RJ. Childhood sleep time and long-term risk for obesity: a 32-year prospective birth cohort study. Pediatrics 2008;122:955–60.
[4] Monasta L, Batty GD, Cattaneo A, Lutje V, Ronfani L, Van Lenthe FJ, et al. Early-life determinants of overweight and obesity: a review of systematic reviews. Obes Rev 2010;11:695–708.
[5] Nielsen LS, Danielsen KV, Sorensen TI. Short sleep duration as a possible cause of obesity: critical analysis of the epidemiological evidence. Obes Rev 2011;12:78–92.
[6] Carter PJ, Taylor BJ, Williams SM, Taylor RW. Longitudinal analysis of sleep in relation to BMI and body fat in children: the FLAME study. BMJ 2011;342:d2712.
[7] Beebe DW, Lewin D, Zeller M, McCabe M, MacLeod K, Daniels SR, et al. Sleep in overweight adolescents: shorter sleep, poorer sleep quality, sleepiness, and sleep-disordered breathing. J Pediatr Psychol 2007;32:69–79.
[8] Spruyt K, Molfese DL, Gozal D. Sleep duration, sleep regularity, body weight, and metabolic homeostasis in school-aged children. Pediatrics 2011;127:e345–52.
[9] Paul IM, Savage JS, Anzman SL, Beiler JS, Marini ME, Stokes JL, et al. Preventing obesity during infancy: a pilot study. Obesity 2011;19:353–61.
[10] Cizza G, Marincola P, Mattingly M, Williams L, Mitler M, Skarulis M, et al. Treatment of obesity with extension of sleep duration: a randomized, prospective, controlled trial. Clin Trials 2010;7:274–85.
[11] Taylor BJ, Heath A-LM, Galland BC, Gray AR, Lawrence J, Sayers R, et al. Prevention of Overweight in Infancy (POI.nz) study: a randomised controlled trial of sleep, food and activity interventions for preventing overweight from birth. BMC Public Health 2011;11:942.
[12] Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc 2005;37:S490–500.
[13] Weiss AR, Johnson NL, Berger NA, Redline S. Validity of activity-based devices to estimate sleep. J Clin Sleep Med 2010;6:336–42.
[14] Sazonov E, Sazonova N, Schuckers S, Neuman M. Activity-based sleep–wake identification in infants. Physiol Meas 2004;25:1291–304.
[15] So K, Buckley P, Adamson TM, Horne RS. Actigraphy correctly predicts sleep behavior in infants who are younger than six months, when compared with polysomnography. Pediatr Res 2005;58:761–5.
[16] Sadeh A, Acebo C, Seifer R, Aytur S, Carskadon MA. Activity-based assessment of sleep–wake patterns during the 1st year of life. Infant Behav Devel 1995;18:329–37.
[17] Gnidovec B, Neubauer D, Zidar J. Actigraphic assessment of sleep–wake rhythm during the first 6 months of life. Clin Neurophysiol 2002;113:1815–21.
[18] Redmond DP, Hegge FW. Observations on the design and specification of a wrist-worn human activity monitoring system. Behav Res Methods Instrum Comput 1985;17:659–69.
[19] Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep wake identification from wrist activity. Sleep 1992;15:461–9.
[20] Anders T, Emde R, Parmelee AE. A manual of standardised terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants. UCLA Brain Information Service. BRI Publications; 1971.
[21] Atlas Task Force, American Sleep Disorders Association. EEG arousals: scoring rules and examples. Sleep 1993;15:174–84.
[22] International Paediatric Work Group on Arousals. The scoring of arousals in healthy term infants (between the ages of 1 and 6 months). J Sleep Res 2005;14:37–41.
[23] Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[24] Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993;46:423–9.
[25] Insana SP, Gozal D, Montgomery-Downs HE. Invalidity of one actigraphy brand for identifying sleep and wake among infants. Sleep Med 2010;11:191–6.
[26] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307–10.
[27] So K, Adamson TM, Horne RS. The use of actigraphy for assessment of the development of sleep/wake patterns in infants during the first 12 months of life. J Sleep Res 2007;16:181–7.
[28] Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep–wake identification: an empirical test of methodological issues. Sleep 1994;17:201–7.
[29] Sadeh A, Acebo C. The role of actigraphy in sleep medicine. Sleep Med Rev 2002;6:113–24.
[30] Montgomery-Downs HE, Crabtree VM, Gozal D. Actigraphic recordings in quantification of periodic leg movements during sleep in children. Sleep Med 2005;6:325–32.
[31] Hyde M, O'Driscoll DM, Binette S, Galang C, Tan SK, Verginis N, et al. Validation of actigraphy for determining sleep and wake in children with sleep disordered breathing. J Sleep Res 2007;16:213–6.
[32] Acebo C, Sadeh A, Seifer R, Tzischinsky O, Hafer A, Carskadon MA. Sleep/wake patterns derived from activity monitoring and maternal report for healthy 1- to 5-year-old children. Sleep 2005;28:1568–77.
[33] Meltzer LJ, Westin AML. A comparison of actigraphy scoring rules used in pediatric research. Sleep Med 2011;12:793–6.
[34] Jean-Louis G, Kripke DF, Mason WJ, Elliott JA, Youngstedt SD. Sleep estimation from wrist movement quantified by different actigraphic modalities. J Neurosci Methods 2001;105:185–91.
[35] Lotjonen J, Korhonen I, Hirvonen K, Eskelinen S, Myllymaki M, Partinen M. Automatic sleep–wake and nap analysis with a new wrist worn online activity monitoring device vivago WristCare. Sleep 2003;26:86–90.