SLEEPJ, 2019, 1–9

doi: 10.1093/sleep/zsz083
Advance Access Publication Date: 3 April 2019
Original Article

Original Article

Actimetry in infant sleep research: an approach to facilitate comparability

Sarah F. Schoch[1], Oskar G. Jenni[2,3], Malcolm Kohler[1] and Salome Kurth[1,*]

[1]Pulmonary Clinic, University Hospital Zurich, Zurich, Switzerland, [2]Child Development Center, University Children's Hospital Zurich, Zurich, Switzerland and [3]Children's Research Center CRC, University Children's Hospital Zurich, Zurich, Switzerland

*Corresponding author. Salome Kurth, Pulmonary Clinic, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland. Email: salome.kurth@usz.ch.

Abstract

Study Objectives: Only standardized objective assessments reliably capture the large variability of sleep behavior in infancy, which is the most pronounced throughout the human lifespan. This is important for clinical practice as well as basic research. Actimetry is a cost-efficient method to objectively estimate infant sleep/wake behavior from limb movements. Nevertheless, the standardization of actimetry-based sleep/wake measures is limited by two factors: the use of different computational approaches and the bias towards measuring only nighttime sleep—neglecting ~20 % of sleep infants obtain during daytime. Thus, we evaluate the comparability of two commonly used actimetry algorithms in infants and propose adjustments to increase comparability.

Methods: We used actimetry in 50 infants for 10 continuous days at ages 3, 6, and 12 months in a longitudinal approach. We analyzed the infants' sleep/wake behaviors by applying two algorithms: Sadeh and Oakley/Respironics. We compared minute-by-minute agreement and Kappa between the two algorithms, as well as the algorithms with sleep/wake measures from a comprehensive 24-hour parent-reported diary.

Results: Agreement between uncorrected algorithms was moderate (77%–84%). By introducing a six-step adjustment, we increased agreement between algorithms (96%–97%) and with the diary. This decreased the difference in estimated sleep behaviors, e.g. *Total Sleep Duration* from 4.5 to 0.2 hours.

Conclusions: These adjustments enhance comparability between infant actimetry studies and the inclusion of parent-reported diaries allows the integration of daytime sleep. Objectively assessed infant sleep that is comparable across different studies supports the establishment of normative developmental trajectories and clinical cutoffs.

Statement of Significance

Actigraphy is a cost-efficient method to estimate sleep/wake behavior from movement. However, generalization of findings in infant sleep research has been limited due to the use of different algorithms for sleep/wake quantification and the primary focus on nighttime sleep. We optimized sleep quantification from actimetry in infants by applying a set of adjustments that overcomes discrepancies between existing algorithms in sleep estimates. This method improves analysis of daytime sleep and leads to increased comparability between studies. In the future, the inclusion of more sensors and a digital diary could lead to development of normative trajectories and enhanced clinical cutoffs.

Key words: actigraphy; pediatrics – infants; scoring; algorithm; actimetry; sleep-wake bias; sleep detection improvement

Submitted: 14 December, 2018; **Revised:** 31 January, 2019

Introduction

Studying the relationship between sleep in early life and later health and behavioral outcomes requires objective and reliable quantitative data. Current practice often relies on parent-reported questionnaires to estimate infants sleep/wake behavior. However, these parent-reported questionnaires are subjective and often disagree with more objective sleep measures, including comprehensive diaries completed by parents across consecutive 24-hour periods, e.g. misjudging sleep duration by >1 hour in young children [1]. Wearables estimate sleep/wake states from arm or leg movement (actimetry) and allow cost-efficient sleep tracking in diverse environments and over long periods of time [2]. Standardized procedures in actimetry studies will facilitate generalization of findings and cross-comparison between studies. Yet, we have to overcome two existing constraints: first, there are no standards for scoring sleep/wake from actimetry. In fact, the comparability of widely used analysis algorithms has not been investigated [3]. Second, it is important to investigate both day- and nighttime sleep in infants as sleep pressure and quality largely depend on the preceding history of day-/nighttime sleep [4]. Certain limitations (e.g. the underestimation of sleep due to external movements from carriage, stroller or bed-sharing, and the overestimation of sleep when immobilized, e.g. baby sling, breastfeeding) have confined most infant actimetry assessments to nocturnal sleep, missing the ~20% of daytime sleep [5].

This study evaluates the comparability of commonly used actimetry methodologies in infants. We compare two approaches and compute their bias to sleep or wake. We then propose adaptations to streamline sleep/wake identification and to quantify infant daytime sleep by integrating 24-hour diary information into the analysis [6]. Increased comparability across actimetry-based sleep estimates reduces sources of variability for ultimately framing sleep–wake patterns and normative sleep in infants.

Methods

Participants

Fifty healthy term-born infants (17 females) were longitudinally assessed with ankle actimetry at age 3 months (i.e. 2.46–3.38 months at assessment start), 6 months (5.42–6.18 months), and 12 months (11.47–12.16 months). The presence of medical conditions (e.g. diseases or lesions of the central nervous system, developmental disabilities, epilepsy, neurologic/metabolic disorders/head injury involving loss of consciousness) and travelling across time zones with >1 hour difference in the 4 weeks prior to assessment served as exclusion criteria. Ethical approval was obtained from the Zurich Ethics Committee (2016-00730) and study procedures were consistent with the declaration of Helsinki. Written parental consent was obtained before enrollment.

Experimental design

Data were collected at each assessment time point for a duration of 10 days (5–16 days) through ankle actimetry and a 24-hour sleep–wake diary. GENEactiv accelerometers (Activinsights Ltd, Kimbolton, UK; 43 × 40 × 13 mm, Micro-Electro-Mechanical Systems sensor, 16 g, 30 Hz frequency; sensitive for ±8 g range at 3.9 mg resolution) were attached on the left ankle with a modified sock or a Tyvek paper strap. Parents were instructed to only remove the actimeter for bathing and to document its removal in the 24-hour diary. The sleep diary was adapted from Werner et al. [1], with parents reporting in 15-minute intervals: sleep (including external movement, e.g. sleeping in the parents arms, stroller *etc.*), wake, feeding, and crying. Parents reported bed times (putting infant to bed in the evening and getting up in the morning) and naps, and marked particular periods of uncertainty (e.g. feeding periods during nighttime). They were instructed to fill out the diary throughout the day. During the assessment, the Brief Infant Sleep Questionnaire (BISQ) was completed [7]. Families received small gifts for the infant (i.e. bottles, baby food) for participation.

Actimetry processing

Actimetric data were extracted as binary files using GENEactiv PC Software (version 3.1), imported into Matlab (R2016b), and converted to activity counts [8], including a 3–11 Hz bandpass filter and signal compression to 15-second bins. Acceleration data from the three axes were combined using a sum of squares. Signal was compressed to one data point per minute by data summation. To identify infant sleep and wake periods, several adjustments were introduced to existing algorithms (Sadeh et al. [9] and Oakley [10]; Figure 1, A). We focused on the most commonly used algorithm in the pediatric literature (Sadeh algorithm as identified in Meltzer et al. [3]) and the most commonly used device (Oakley algorithm used with Respironics devices). We implemented adjustments to algorithms that were based on frequently applied procedures (threshold, rescoring, smoothing [11–13]). An additional adjustment was used to counteract a bias towards overemphasizing sleep or wake. The order of adjustments was tailored to first adapt the algorithm (threshold and bias factor) and then adapt the scoring of sleep and wake (corrections based on sleep diary, Webster rescoring [14], smoothing). Further, the order of adjustments was chosen to prevent that adjustments interact (e.g. first the correction for external movement was completed, then smoothing was applied).

The following adjustments were performed: The first adjustment include the change of threshold. For the Oakley algorithm, this refers to the identification of the value serving as a threshold to distinguish sleep from wake. While generally a threshold of 20, 40, or 80 is used, we replaced the threshold with mean activity of the full recording*0.888 (similar to the auto-threshold setting). This threshold has been shown to work best in infants older than 2 months [11]. In contrast, the Sadeh algorithm applies a threshold to distinguish between low- and high-activity epochs. This threshold is originally set 100, which we replaced consequently with mean activity of the full recording*0.888. The second adjustment was the introduction of factor based on mean activity of each recording [12], with the aim of counteracting the strong bias to either sleep or wake of each algorithm. This factor was added (Sadeh)/subtracted (Oakley) from the computed activity (Figure 1, B). The third adjustment included the replacement of periods when the actimeter was not worn. These periods were identified with the parent-reported 24-hour diary (Figure 1, C). This adjustment step counteracted incorrect scoring of sleep originating from a lack of activity. The fourth

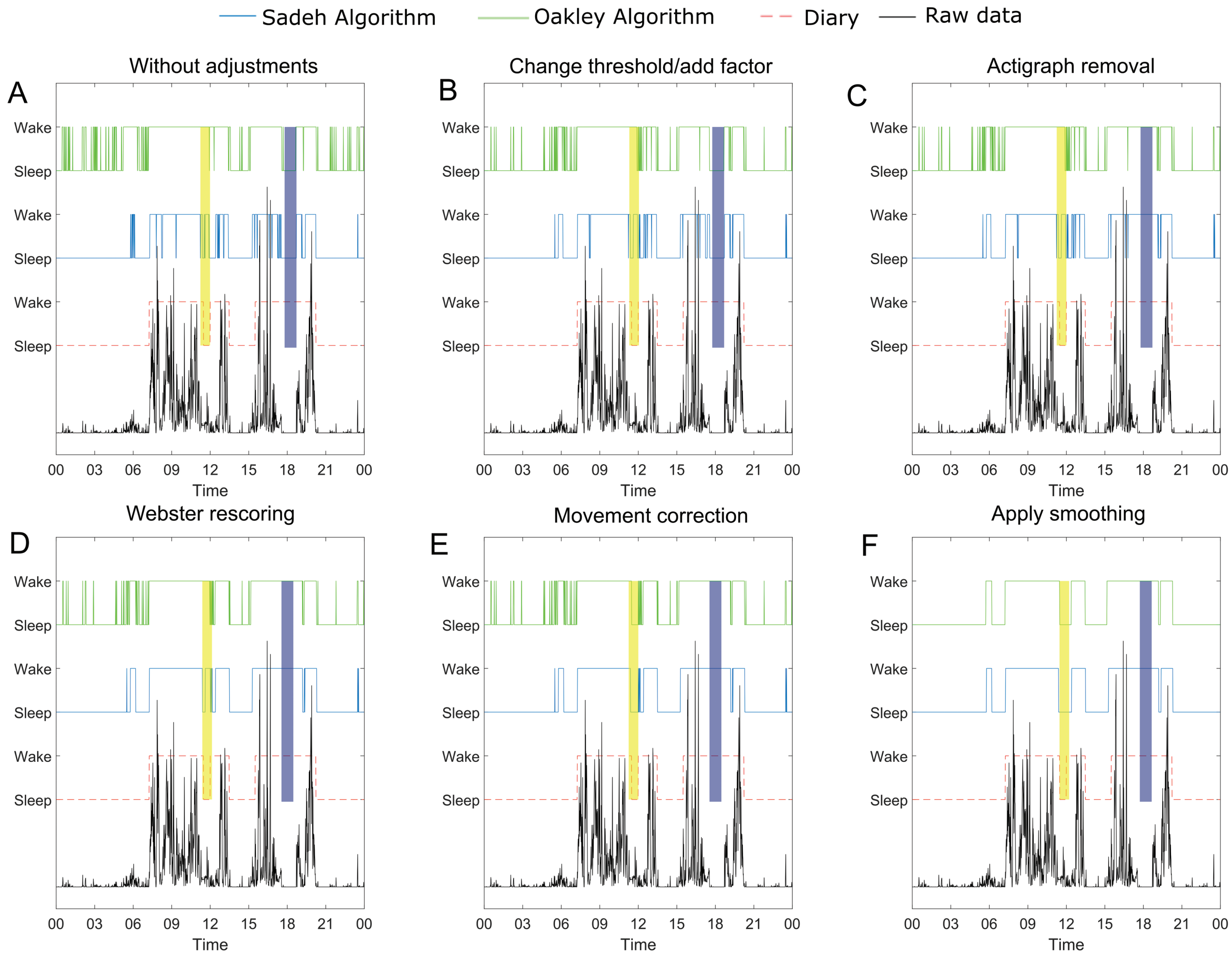

Figure 1. Stepwise processing adjustments. Typical 24-hour actimetric profile from a representative participant (age 12 months). Raw data (black) and the scorings from 24-hour diary (red), Sadeh (blue), and Oakley (green) are presented. Wake is shown on top and sleep at the bottom of each scoring item. Stepwise adjustments are presented in order of processing: (A) raw data without adjustments; (B) altered threshold and added factor reducing wake/sleep bias; (C) rescoring of actimeter removal with 24-hour diary information; (D) rescoring by Webster; (E) rescoring of sleep with external movements; and (F) smoothing of short wake periods (<5 minutes) during sleep. Yellow shading indicates periods of sleep with reported external movement. Blue shading illustrates periods with actimeter removal.

adjustment was implemented to rescore data using the strict criteria by Webster et al. [14]. This corrects mis-scoring of sleep by addressing short periods of inactivity during wake (short periods [6–10 minutes] of sleep surrounded by periods of wake and the first 1–4 minutes of sleep are rescored wake, Figure 1, D). The fifth adjustment was the correction of sleep periods with known external movement, as verified with the diary (Figure 1, E). Finally, the sixth adjustment included data smoothing. Wake periods shorter than 5 minutes were removed if they were surrounded by sleep periods. Because infants generally show higher movement activity during sleep (twitches) than older children [15, 16], smoothing ensures that sleep periods with movement are still scored as sleep (Figure 1, F).

In order to reduce error caused by external factors, 24-hour days were excluded for the calculation of sleep variables if (1) the actimeter was removed for >3 hours (in 22.2% of all data including first and last days of the assessment, and interruptions due to sickness), (2) the infant was sick but the overall assessment was continued (4.2%), or (3) the assessment took place during the switch to/from daylight savings (0.2%). These criteria resulted in the following data included in final analysis: mean assessment duration of 8.6 ± 1.65 days at age 3 months (whereby 3 days was the minimum assessment duration and 13 days, the maximum), correspondingly 8.0 ± 1.95 days included at 6 months (2–11 days), and 7.9 ± 1.71 days included at 12 months (3–10 days).

From the resulting matrix containing a minute-by-minute scoring of either sleep or wake, sleep variables of interest were computed: *Total Sleep Duration*, *Day-to-Day Sleep Variability*, *% Night Sleep*, and *Fragmentation*. *Total Sleep Duration* (hour) sums the time scored as sleep within 24 hours (starting at clock time 0:01). *Day-to-Day Sleep Variability* (hour) is the *SD* of the *Total Sleep Duration* across all included assessment days. *% Night Sleep* indicates the relative proportion of nighttime sleep (i.e. within clock time 19:00–07:00) as a percentage of *Total Sleep Duration*. *Fragmentation* (awakenings/hour) calculates the number of awakenings per hour during nighttime sleep (based on individual infant bedtimes reported by parents). Awakenings were scored separate when divided by at least 10 minutes of sleep. BISQ total sleep duration was calculated by adding reported day and night sleep duration (rounded to 15 minutes; mean was used when time range was reported). Nine BISQ assessments were excluded due to incomplete data. To calculate the agreement between actigraphy and the 24-hour diary, the diary was transformed to a minute-by-minute scoring resolution. Feeding periods during sleep were scored as wake.

Statistical analysis

We used R (version 3.5.0) and R Studio (version 1.1.463) for statistical analyses. Linear mixed-effect models were estimated using restricted maximum likelihood to analyze changes resulting from adjustments using the R-packages *lmer* [17] and *lmertest* [18]. The covariate assessment time point was included as a logarithmic function of age (log[age]). We chose this logarithmic function to account for the flattening of effects with age (larger effects between 3 and 6 months than between 6 and 12 months). All models included effects of adjustment, infant age, and their interaction. To compare whether random effects of time point and adjustments improve model fit, we compared one model combining both random effects with two separate models containing random effects of either time point or adjustment. The random effects were only included in the final model if it significantly improved the model fit with most weight given to the Bayesian Information Criterion (BIC; Supplementary Tables 1–5, selected model highlighted in bold).

We calculated agreement between two measures as percentage of 1-minute periods scoring the same state (i.e. sleep or wake) and additionally using Cohen's Kappa [19]. Bias was calculated as the difference (minute) where one algorithm scored sleep and the other wake. We used Bland–Altman statistics to investigate whether the algorithms calculated similar estimates for sleep variables (package *BlandAltmanLeh*). Further, we tested the stability across age by investigating whether large differences between algorithms at age 3 months are also associated with large differences at older age. Accordingly, we performed correlations between time points with the difference measure resulting from the Bland–Altman statistics. A two-sided significance level of p <0.05 was used.

Results

Agreement between algorithms (Sadeh algorithm–Oakley algorithm)

We compared the agreement between the algorithms with and without adjustments. Without adjustment, algorithms show moderate agreement in scoring sleep or wake (77%–84%, κ = 0.50–0.68; Table 1). Agreement was significantly improved by introducing the six-step adjustment (96%–97%, κ = 0.91–0.95, $t_{(274.63)} = 23.35$, $p < 0.0001$; Supplementary Table 1). The largest disagreement was observed in actimetry data from infants aged 3 months ($t_{(247)} = 14.44$, $p < 0.0001$). The largest improvement in agreement occurred at age 3 months (interaction age * improvements, $t_{(247)} = -7.63$, $p < 0.0001$). The improved agreement mainly results from threshold adaptation and adding the factor against bias (~5%) as well as smoothing (~2%).

Agreement between algorithms and 24-hour diary (Sadeh algorithm–24-hour-diary, Oakley algorithm–24-hour-diary)

We compared the scorings of both algorithms, with and without adjustments, with the parental-reported 24-hour diary. Both algorithms showed medium agreement with the 24-hour diary without adjustments (75%–85%, Sadeh vs Diary κ = 0.51–0.70, Oakley vs Diary κ = 0.5–0.68; Table 1). Adjustments increased agreement to up to 93% (86%–93%, Sadeh vs Diary κ = 0.72–0.86, Oakley vs Diary κ = 0.71–0.86, $t_{(494.42)} = 13.93$, $p < 0.0001$; Supplementary Table 2). Lower agreement was seen for 3-month olds compared to 6- and 12-month olds ($t_{(495)} = 12.19$, $p < 0.0001$). The observed interaction between age and improvement ($t_{(495)} = -3.31$, $p = 0.001$) indicates that adjustments lead to greater improvements particularly in the youngest age group. There was no significant effect of algorithm (Sadeh vs Oakley $t_{(1,495)} = 1.90$, $p = 0.06$) and no interaction of type of algorithm and age ($t_{(495)} = -0.32$, $p = 0.75$). A small interaction was observed between algorithm and amount of improvements, with the Oakley algorithm showing increased improvements due to the adjustments ($t_{(495)} = -2.02$, $p = 0.04$). At 3 months, adjusting for movement during sleep greatly improved the agreement (~4.5%), which was less pronounced for 6 and 12 months, respectively (~1.5%–3 %). The opposite was seen for adjustments for actimeter removal, which occurred less at 3 months (0.68%) than at 6 and 12 months (~1.25%–2%).

Bias towards sleep or wake (Sadeh algorithm–Oakley algorithm, algorithms–24-hour-diary)

Each algorithm had a scoring bias for a specific state: 200–300 minutes per day were scored as sleep by the Sadeh algorithm and wake

Table 1. Agreement rates with and without adjustment steps

Age	No adjustments	Change threshold	Add factor	Actigraph removal	Rescoring Webster	External movements	Smoothing
Sadeh–Oakley							
3 months	77.36 ± 3.97	83.21 ± 3.75	91.56 ± 1.59	91.64 ± 1.59	92.96 ± 1.34	93.56 ± 1.38	95.76 ± 1.00
6 months	83.65 ± 3.67	87.65 ± 3.68	93.14 ± 1.34	93.23 ± 1.33	93.43 ± 1.13	94.89 ± 1.16	96.79 ± 0.80
12 months	83.97 ± 3.07	87.99 ± 3.12	93.66 ± 1.32	93.78 ± 1.32	94.66 ± 1.34	94.88 ± 1.36	97.43 ± 0.83
Sadeh–Diary							
3 months	76.39 ± 6.05	75.25 ± 6.11	79.55 ± 5.19	80.22 ± 5.16	81.65 ± 5.17	86.22 ± 5.17	86.36 ± 5.20
6 months	82.26 ± 4.93	81.18 ± 4.92	84.42 ± 3.59	85.67 ± 3.60	86.86 ± 3.59	89.60 ± 3.21	89.69 ± 3.26
12 months	85.40 ± 6.23	84.30 ± 6.12	88.19 ± 5.11	90.21 ± 2.87	91.55 ± 2.66	93.07 ± 2.42	93.23 ± 2.44
Oakley–Diary							
3 months	75.14 ± 4.57	77.17 ± 4.83	77.22 ± 4.86	77.91 ± 4.81	79.24 ± 7.40	84.05 ± 4.95	85.73 ± 5.08
6 months	80.77 ± 3.34	82.77 ± 3.53	82.77 ± 3.55	84.00 ± 3.56	84.76 ± 3.41	87.85 ± 3.21	89.45 ± 3.25
12 months	84.15 ± 5.04	86.34 ± 5.11	86.36 ± 5.13	88.38 ± 2.83	89.08 ± 2.86	90.76 ± 2.72	92.99 ± 2.50

Agreement rates as % agreement (averaged over participants over measurement days). Agreements are shown between the two algorithms and between each algorithm and sleep diaries filled out by the parents. Means ± *SD* is shown.

by the Oakley algorithm (Figure 2). This bias was significantly reduced by the adjustments ($t_{(282.49)} = -27.34$, $p < 0.0001$; Supplementary Table 3). Particularly 3-month-olds' scorings showed increased bias in comparison with the older infants' scorings ($t_{(247)} = -13.35$, $p < 0.0001$), but bias decreased most through our adjustments at that age ($t_{(1,247)} = 11.45$, $p < 0.0001$). Similar bias was observed when compared to the 24-hour diary: the Sadeh algorithm scored more sleep than reported in the 24-hour diary. This bias decreased through the adjustments ($t_{(198)} = -6.38$, $p < 0.0001$). The bias was stronger with lower age ($t_{(100.58)} = -2.79$, $p = 0.006$) but showed no interaction with age ($t_{(198)} = 1.50$, $p = 0.13$). The Oakley algorithm scored more wake compared to the sleep 24-hour diary. This bias was significantly reduced by our adjustments ($t_{(81.81)} = 11.68$, $p < 0.0001$). There was an age effect ($t_{(90.08)} = 3.50$, $p = 0.0007$), with the largest improvements at 3 months ($t_{(198)} = -5.75$, $p < 0.0001$).

Sleep/wake behavior estimation (Sadeh algorithm–Oakley algorithm)

To estimate differences in the sleep parameters, we calculated Bland–Altman statistics of each parameter without and with adjustments (Table 2 and Figures 3–6). Without adjustments, there was a bias in the variables *Sleep Duration*, *% Night Sleep*, and *Fragmentation*, as shown by data points instead of being centered around 0 (no bias), they were centered around, e.g., 4.2 hours for *Sleep Duration* (Figure 3). Bias for each age group is shown in Table 2, e.g. 5.45 hours at age 3 months (previously this approach was used with a bias definition exceeding ± 0.5 hour) [1]. This bias was reduced by our adjustments to, *e.g.*, –0.01 hour in *Sleep Duration* at 3 months. The only variable showing low bias (mean < 0.5 hour) already without adjustments was *Day-to-Day Sleep Variability*. Taken together, we show that infant actimetry-based detection of sleep/wake variables can be improved by six-steps of adjustments.

We then tested whether differences between algorithms were stable across age. We analyzed whether differences between algorithms in sleep variable estimation were correlated between time points (i.e., 3 vs 6 resp. 12 months, Table 3). Thereby, strong positive correlations indicate a low age-effect, and for instance reveal that large differences in sleep estimates at age 3 months also show large differences at age 6 months. Generally, correlations were stronger for unadjusted data, but did not reach statistical significance.

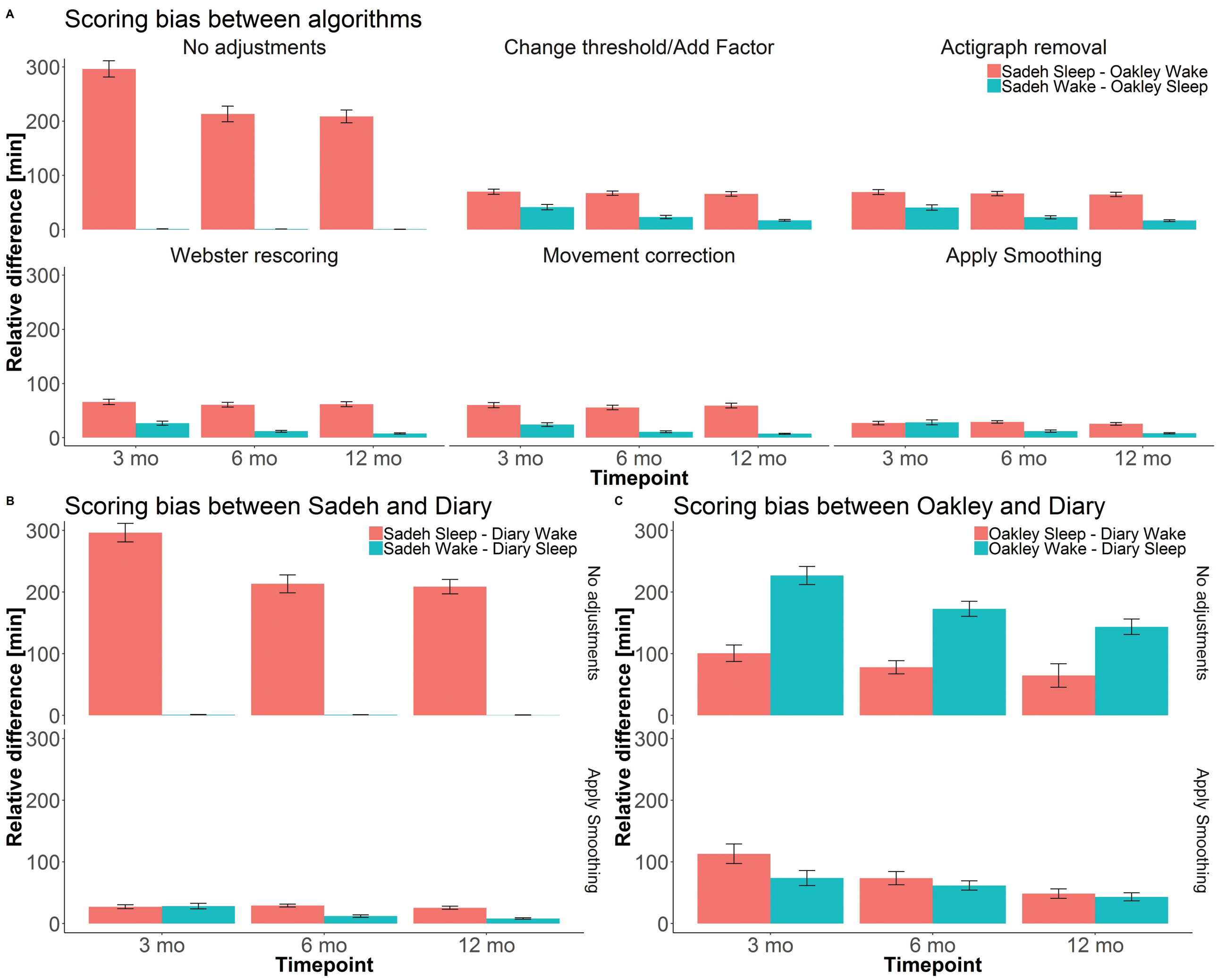

Figure 2. Sleep/wake bias of scoring algorithms and 24-hour diary reported by parents. Scoring bias shows disagreement of scoring as sleep or wake (sum of minutes within 24 hours, error bars represent 95% confidence interval). (A) Scoring bias is depicted without adjustments and for each adjustment step. (B) Scoring bias of the Sadeh algorithm and compared to the 24-hour diary is shown without adjustments and including all adjustments. (C) Bias of the Oakley algorithm compared to the diary is shown without adjustments and including all adjustments.

Table 2. Differences in sleep parameter estimates with and without adjustments estimated by Bland–Altman scores

	Sleep Duration (hour)		Day-to-Day Sleep Variability (hour)		% Night Sleep		Fragmentation (/hour)	
Age	No adjustments	After adjustments	No adjustments	After adjustments	No adjustments	After adjustments	No adjustments	After adjustments
3 months	5.45 ± 1.90	−0.01 ± 0.90	−0.09 ± 0.70	0.01 ± 0.22	−6.99 ± 5.74	0.11 ± 1.52	−0.65 ± 0.37	−0.04 ± 0.09
6 months	3.96 ± 1.69	0.33 ± 0.41	−0.06 ± 0.81	0.00 ± 0.22	−6.64 ± 6.33	−0.40 ± 1.08	−0.78 ± 0.35	−0.06 ± 0.12
12 months	4.08 ± 1.46	0.32 ± 0.37	−0.09 ± 0.55	−0.09 ± 0.20	−4.74 ± 5.60	3.35 ± 1.34	−0.92 ± 0.32	−0.17 ± 0.08

Means ± critical difference is shown. Means show general bias of one measure over the other. Critical difference show 95% differences in estimates.

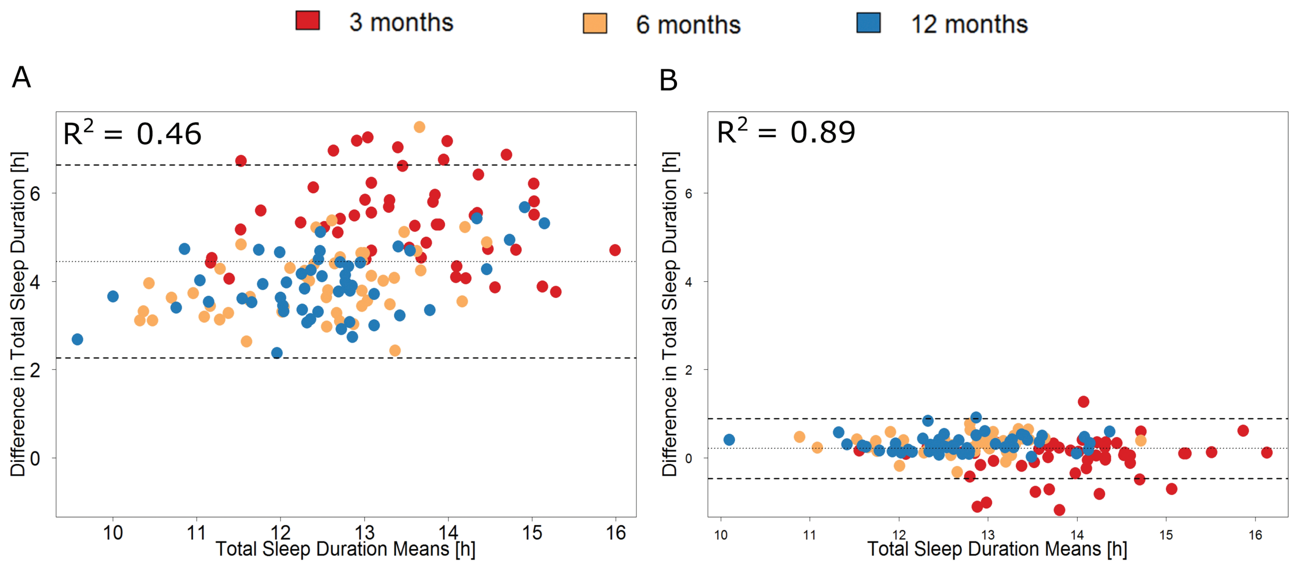

Figure 3. Bland–Altman plots of *Total Sleep Duration* estimates from Sadeh and Oakley algorithm. Each infant is represented by three dots indicating the age group by color. (A) Without adjustments, the difference in *Total Sleep Duration* is >4 hours with a critical difference of 2.18 hours, indicating that without adjustments *Total Sleep Duration* estimates from the Sadeh algorithm are 2–7 hours above estimates from the Oakley algorithm. (B) With six-step adjustments, the difference in *Total Sleep Duration* is lowered to ~ 0 hour with a critical difference of 0.68 hour.

Comparison with BISQ (Sadeh algorithm–questionnaire, Oakley algorithm–questionnaire)

We found a generally large deviation from parental questionnaire data compared to actimetry data in *Total Sleep Duration*, as indicated by a critical difference of 3.19 hours. This includes both, under- and overestimating of the objective estimates by parent's estimates (95%; see Supplementary Figures 1 and 2). For example, parents reporting their infants' sleep duration to be 13.5 hours revealed objectively measured infant sleep duration between 11.86 and 14.6 hours. However, there was no systematic bias (i.e. either under- or overestimating) *Total Sleep Duration* of their infants.

Discussion

Only standardized objective assessments reliably capture the large variability of sleep behavior in infancy, which is the most pronounced during the human lifespan [5]. When polysomnographic recording is not a feasible approach to measure sleep in real-life settings with large populations, actimetry can transform movement counts into objective sleep estimates. We applied a six-step set of adjustments to actimetry-based sleep estimation designed for infants, with the goal to overcome discrepancies in sleep estimates between existing scoring algorithms [9, 10]. The use of 24-hour diaries minimizes signal miscomputation through external factors and improves the analysis of daytime sleep. These methods will help to extend reference values based on parental reports [5] or meta-analysis based on different devices [20].

Adjustments reduced disagreement between algorithms from 16%–22% to 3%–4%. Both algorithms showed a bias when compared to the 24-hour diary and to the other algorithm: the Sadeh algorithm was biased towards sleep and the Oakley algorithm was biased towards wake. Both biases were significantly reduced by the adjustments. Such standardization is of great importance for computation of sleep variables. For example, without adjustments, *Total Sleep Duration* deviates up to 7 hours depending on the algorithm used, with higher sleep duration estimates when using the Sadeh algorithm compared to the Oakley algorithm. After adjustments, these estimates vary less than 1 hour. Importantly, this also increased the correlation, meaning that the infants who overall showed the highest sleep duration as calculated from one algorithm also are estimated to have a high sleep duration with the other algorithm. Similar effects were seen for parameters such as *% Night Sleep* and *Fragmentation*. Only *Day-to-Day Sleep Variability* showed no bias without adjustments, but even for this parameter correlation could be improved drastically.

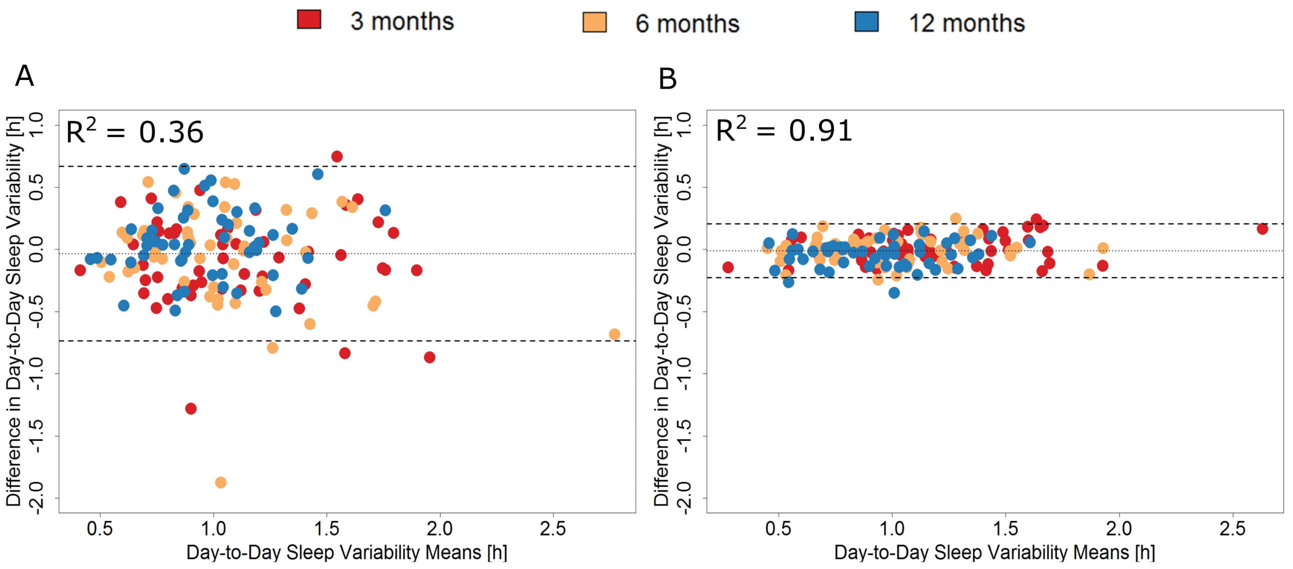

Figure 4. Bland–Altman plots showing difference in *Day-to-Day Sleep Variability* between scoring based on Sadeh and Oakley algorithms. Each infant is represented by three dots indicating the age group by color. (A) Without adjustments, the difference in *Day-to-Day Sleep Variability* is ~0 hour with a critical difference of 0.7 hour. (B) With six-step adjustments, the difference in *Day-to-Day Sleep Variability* estimate is ~ 0 hour with a critical difference of 0.22 hour.

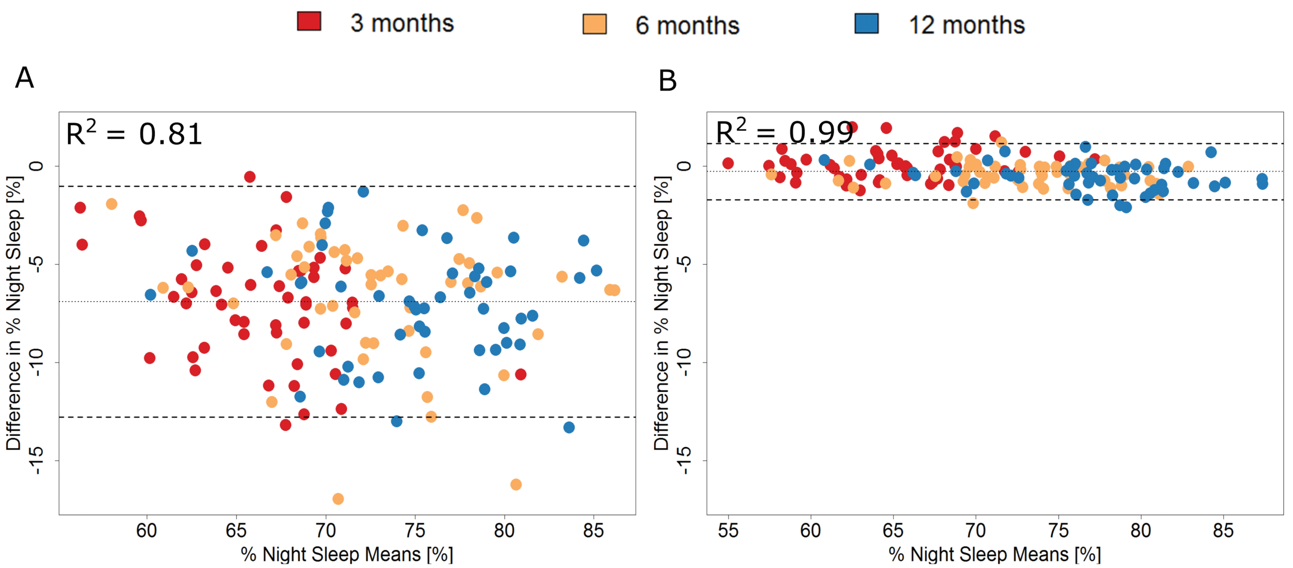

Figure 5. Bland–Altman plots showing difference in *% Night Sleep* between scoring based on Sadeh and Oakley algorithm. Each infant is represented by three dots indicating the age group by color. (A) Without adjustments, the mean difference in *% Night Sleep* estimate is ~7% with a critical difference of 5.87%. (B) With six-step adjustments, the mean difference in *% Night Sleep* estimate is ~0% with a critical difference of 1.43%.

We tested the stability across age of the approach. Effects were stable from 3 to 6 months of age in *Fragmentation*, yet no systematic effect was found in other variables. This suggests that factors compromising the algorithm output are not stable traits of the infant, e.g. small differences in activity between wake and sleep might account for differences at 3 months but does not persist through older age.

We also identified age-specific effects that affect actimetry outcomes. Scoring agreement generally increases with age. We hypothesize that this is primarily due to increased motor activity during wake as part of motor development. Another contributing factor may be the reduction of night-feedings as infants grow older. In the 24-hour diary, feeding was assigned to wake, but transitions of feeding to sleep may be blurry and this might contribute to differences between diaries and scoring. Nonetheless, benefits of algorithm adjustments still prevailed in the older infants (with fewer feedings). Additionally, external movement during sleep in very young infants can lead to mis-scoring of up to 1 hour. This was corrected by introducing our adjustments. As children get older and transit to nap only in a bed, this adjustment may become redundant. Furthermore, with increasing age, removal time of the actimeters increased (e.g. removal by child or other infants, longer periods of bathing/water activities), which led to mis-scoring of up to 30 minutes. Completing the 24-hour diary remains important for the reliable detection and correction of such incidents.

Information from the 24-hour diary support the integration of the ~20% of infant sleep that occurs during daytime. Daytime naps are often missed in traditional analyses, but they reflect

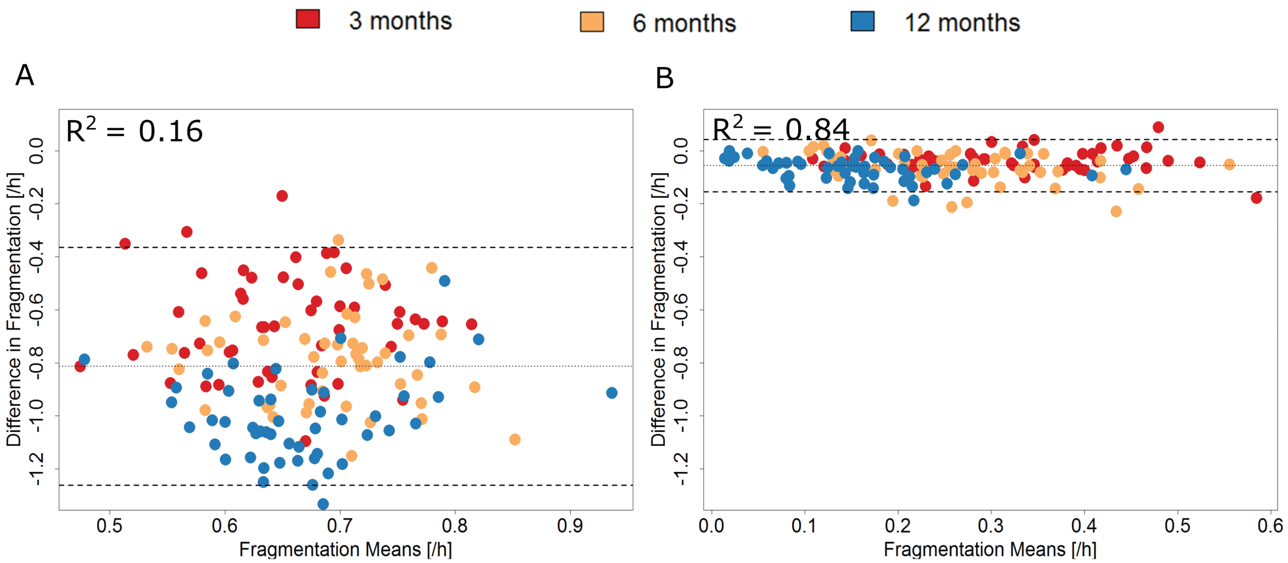

Figure 6. Bland–Altman plots showing difference in *Fragmentation* between scoring based on Sadeh and Oakley algorithm. Each infant is represented by three dots indicating the age group by color. (A) Without adjustments, the mean difference in *Fragmentation* estimate is ~0.8% with a critical difference of 0.45. (B) With six-step adjustments, the mean difference in *Fragmentation* estimate is ~0 with a critical difference of 0.1%.

Table 3. Pearson's correlations testing age effects and comparison of age groups are based on the differences between the two algorithms in sleep variables

	Sleep Duration (hour)		Day-to-Day Sleep Variability (hour)		% Night Sleep		Fragmentation (/hour)	
Correlation *r*	No adjustments	After adjustments	No adjustments	After adjustments	No adjustments	After adjustments	No adjustments	After adjustments
3–6 months	0.31	0.09	-0.21	0.009	0.31	0.001	**0.49**	0.05
3–12 months	0.09	0.13	-0.13	0.23	0.20	0.21	0.26	0.07

Bold value is significant after FDR correction.

the important build-up of sleep pressure and the neurophysiological capacity of children to increase consolidated waking bouts [21]. Our approach circumvents these difficulties by integrating complementary information from a 24-hour sleep diary. Although our semi-automated integration requires time investment of study participants and researchers, it greatly improves data reliability and allows comparison across studies. We suggest to integrate digital diaries (i.e. sleep tracking apps) linked to actimetry input for future studies. Parents should be given the opportunity to confirm sleep periods or reject faulty ones electronically. Additional computational corrections can be introduced to (1) distinguish between movements of the infant vs external movements or (2) automatically detect periods where the actimeter is not worn. This requires the integration of new sensors such as heart rate or skin temperature. Such sensors could also distinguish quiet wakefulness from sleep, which cannot be achieved with acceleration only.

Limitations

This investigation aimed at quantifying infant sleep in real-life settings and did thus not compare actimetry or 24-hour diary data with simultaneously assessed polysomnography. Polysomnography is the current gold-standard objective sleep measure, yet its recording was not feasible in the frame of the current research (50 infants, multiple recordings throughout the first year of life). A further caveat of this research is that the specific infant Sadeh algorithm is validated only against observer rating [9], in contrast to a similar algorithm, which was validated against polysomnography in young children [22, 23]. Yet, the Oakley algorithm was validated against polysomnography in infants [11]. Remaining validations are clearly needed, and we anticipate that the here proposed analytical adjustments will further increase agreement between actimetric and electrophysiological measures of sleep in infants. Principally, we investigated infant data and some of the adjustments might be specific to this age only, while others might even transfer to older age groups. A systematic investigation in older age groups would identify which adjustments support data processing and interpretation.

In conclusion, we present adjustments to standardize actimetric sleep/wake scoring for nighttime and daytime sleep. Applying these adjustments increases the reliability of measured infant sleep variables.

Supplementary material

Supplementary material is available at *SLEEP* online.

Funding

This work was supported by the Clinical Research Priority Program Sleep and Health of the University of Zurich (Clinical Research Priority Program Sleep and Health; Forschungskredit FK-18-047; Medical Faculty; to SK), the Swiss National Science Foundation (P0ZHP1-178697 to SFS), the Medical Faculty of the University of Zurich and the Forschungskredit of the University of Zurich, grant no. FK-18-047 (to SK).

Deposit of data and scripts: doi: 10.17605/OSF.IO/6S3PW

A preprint of this article was deposited at bioRxiv: https://doi.org/10.1101/494427

Conflict of interest statement. None declared.

References

1. Werner H, *et al*. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns: recommendations for clinical and research practice. *J Sleep Res*. 2008;**17**:127–128.
2. Ancoli-Israel S, *et al*. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003;**26**(3):342–392.
3. Meltzer LJ, *et al*. Use of actigraphy for assessment in pediatric sleep research. *Sleep Med Rev*. 2012;**16**(5):463–475.
4. Borbély AA. A two process model of sleep regulation. *Hum Neurobiol*. 1982;**1**(3):195–204.
5. Iglowstein I, *et al*. Sleep duration from infancy to adolescence: reference values and generational trends. *Pediatrics*. 2003;**111**(2):302–307.
6. Tsai SY, *et al*. Actigraphy as a measure of activity and sleep for infants: a methodologic study. *Arch Pediatr Adolesc Med*. 2010;**164**(11):1071–1072.
7. Sadeh A. A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample. *Pediatrics*. 2004;**113**(6):e570–e577.
8. te Lindert BH, *et al*. Sleep estimates using microelectromechanical systems (MEMS). *Sleep*. 2013;**36**(5):781–789.
9. Sadeh A, *et al*. Activity-based assessment of sleep-wake patterns during the 1st year of life. *Infant Behav Dev*. 1995;**18**(3):329–337.
10. Oakley N. *Validation With Polysomnography of the Sleepwatch Sleep/Wake Scoring Algorithm Used by the Actiwatch Activity Monitoring System*. Bend, OR: Mini Mitter, Cambridge Neurotechnology; 1997.
11. So K, *et al*. Actigraphy correctly predicts sleep behavior in infants who are younger than six months, when compared with polysomnography. *Pediatr Res*. 2005;**58**(4):761–765.
12. Tilmanne J, *et al*. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *J Sleep Res*. 2009;**18**(1):85–98.
13. Bélanger MÈ, *et al*. Validating actigraphy as a measure of sleep for preschool children. *J Clin Sleep Med*. 2013;**9**(7):701–706.
14. Webster JB, *et al*. An activity-based sleep monitor system for ambulatory use. *Sleep*. 1982;**5**(4):389–399.
15. Tonetti L, *et al*. Actigraphic motor activity during sleep from infancy to adulthood. *Chronobiol Int*. 2017;**34**(2):246–253.
16. Blumberg MS, *et al*. Twitching in sensorimotor development from sleeping rats to robots. *Curr Biol*. 2013;**23**(12):R532–R537.
17. Bates D, *et al*. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;**67**:1–48.
18. Kuznetsova A, *et al*. lmerTest package: tests in linear mixed effects models. *J Stat Softw*. 2017;**82**(13):1–26.
19. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;**70**(4):213–220.
20. Galland BC, *et al*. Normal sleep patterns in infants and children: a systematic review of observational studies. *Sleep Med Rev*. 2012;**16**(3):213–222.
21. Kurth S, *et al*. Development of nap neurophysiology: preliminary insights into sleep regulation in early childhood. *J Sleep Res*. 2016;**25**(6):646–654.
22. Sadeh A, *et al*. Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. *Pediatrics*. 1991;**87**(4):494–499.
23. Sadeh A, *et al*. Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications. *J Ambulatory Monitor*. 1989;**2**(3):209–216.