SLEEP, 2023, **46**, 1–14
https://doi.org/10.1093/sleep/zsad160
Advance access publication 6 June 2023
Original Article

Original Article

A longitudinal study of infant 24-hour sleep: comparisons of sleep diary and accelerometer with different algorithms

Tiange Liu[1] and Sara E. Benjamin-Neelon[1,2,*]

[1]Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and
[2]Department of International Health, Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*Corresponding author. Sara E. Benjamin-Neelon, PhD, JD, MPH, 615 N. Wolfe Street, Baltimore, Maryland 21205, USA. Email: sara.neelon@jhu.edu.

Abstract

Study Objectives: To longitudinally compare sleep/wake identification and sleep parameter estimation from sleep diaries to accelerometers using different algorithms and epoch lengths in infants.

Methods: Mothers and other caregivers from the Nurture study (southeastern United States, 2013–2018) reported infants' 24-hour sleep in sleep diaries for 4 continuous days, while infants concurrently wore accelerometers on the left ankle at 3, 6, 9, and 12 months of age. We applied the Sadeh, Sadeh Infant, Cole, and Count-scaled algorithm to accelerometer data at 15 and 60 seconds epochs. For sleep/wake identification, we assessed agreement by calculating epoch-by-epoch percent agreement and kappas. We derived sleep parameters from sleep diaries and accelerometers separately and evaluated agreement using Bland–Altman plots. We estimated longitudinal trajectories of sleep parameters using marginal linear and Poisson regressions with generalized estimation equation estimation.

Results: Among the 477 infants, 66.2% were black and 49.5% were female. Agreement for sleep/wake identification varied by epoch length and algorithm. Relative to sleep diaries, we observed similar nighttime sleep offset, onset, and total nighttime sleep duration from accelerometers regardless of algorithm and epoch length. However, accelerometers consistently estimated about 1 less nap per day using the 15 seconds epoch, 70 and 50 minutes' shorter nap duration per day using the 15 and 60 seconds epoch, respectively; but accelerometers estimated over 3 times more wake after nighttime sleep onset (WASO) per night. Some consistent sleep parameter trajectories from 3 to 12 months from accelerometers and sleep diaries included fewer naps and WASOs, shorter total daytime sleep, longer total nighttime sleep, and higher nighttime sleep efficiency.

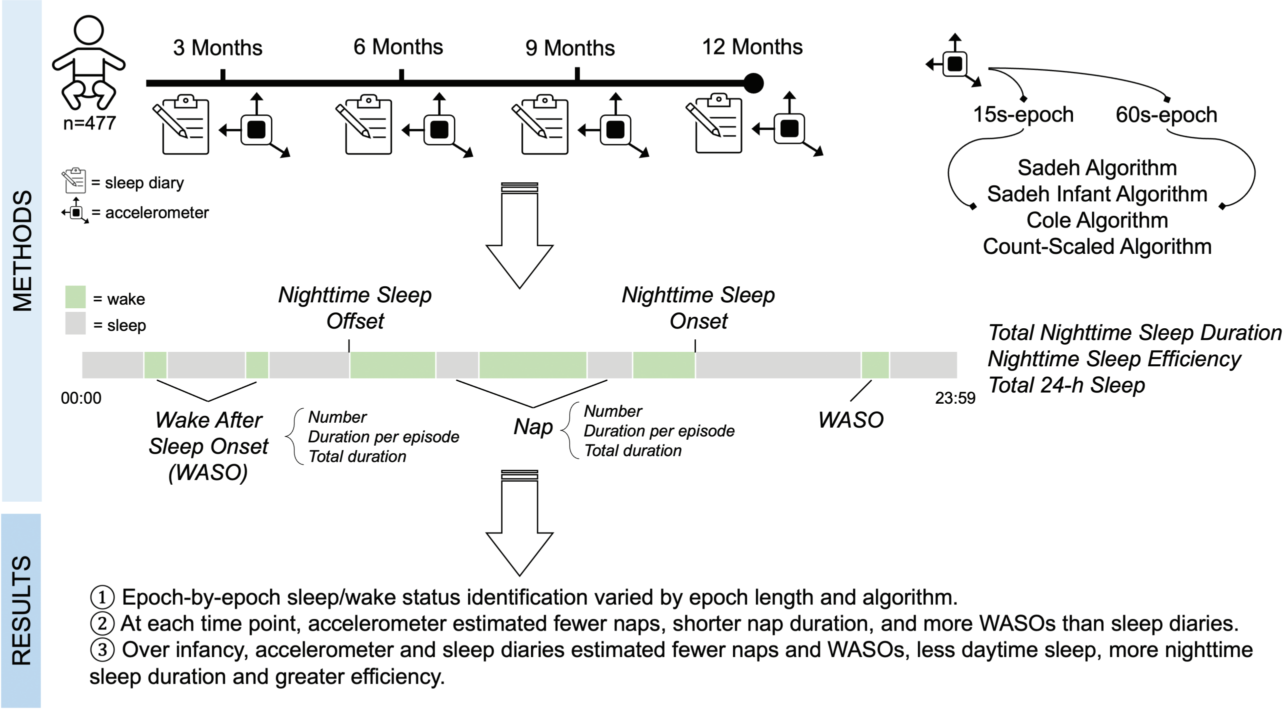
Conclusions: Although there is no perfect measure of sleep in infancy, our findings suggest that a combination of accelerometer and diary may be needed to adequately measure infant sleep.

Key words: infant; accelerometer; sleep diary; algorithm; sleep/wake identification

---

Submitted for publication: December 13, 2022; **Revised:** May 22, 2023

Graphical Abstract



① Epoch-by-epoch sleep/wake status identification varied by epoch length and algorithm.
② At each time point, accelerometer estimated fewer naps, shorter nap duration, and more WASOs than sleep diaries.
③ Over infancy, accelerometer and sleep diaries estimated fewer naps and WASOs, less daytime sleep, more nighttime sleep duration and greater efficiency.

Statement of Significance

Sleep is a key component of infant health and development, but few studies have assessed longitudinal and more objectively measured sleep in a large sample of racially diverse infants. We assessed absolute values and longitudinal changes in 24-hour sleep using caregiver-reported diaries and accelerometers in over 400 infants. We found that relative to diaries, accelerometers underestimated naps, and overestimated night wakings. However, diaries and accelerometers were in agreement for other sleep parameters, such as duration and efficiency of nighttime sleep. We found general agreement for 12-month trajectories, including decreasing naps, night wakings, and daytime sleep, and higher nighttime sleep efficiency. Accelerometer data using the Sadeh Infant algorithm had greater concordance with sleep diaries than the Sadeh, Cole, and Count-scaled algorithms.

Introduction

Infants experience major changes in sleep behaviors as a result of rapid neurobehavioral development and cognitive functioning during the first year of life. As infants age, they are expected to spend more time awake than asleep, have more consolidated sleep periods, and fall asleep and resume sleep with little regulation by caregivers [1]. Yet, what constitutes healthy, normal sleep for infants is highly dependent on biological and sociocultural factors [2, 3]. Unfavorable sleep habits during infancy may persist into childhood [4, 5] and contribute to adverse metabolic, psychological, and developmental consequences in childhood and beyond [6–8]. However, the need for sleep varies both between infants of the same age and within an individual infant throughout the first year. Nevertheless, reference values of sleep parameters are still useful to help parents, clinicians, and researchers identify sleep problems in infancy. A 2018 systematic review summarized reference values and changes in sleep parameters in infants reported by 47 research articles conducted globally over the past 20 years [9]. The authors confirmed the existence of discrepancies across studies in reference values of sleep parameters and emphasized the importance of methodological differences that need to be considered when interpreting findings between studies.

Different measurement methods used by previous studies may have significant impacts on the estimation of infant sleep [10–18]. Parent report via sleep diary, questionnaire, or interview is the most commonly used method partially due to its cost-effectiveness and less labor-intensive nature [9]. However, parent report is known to be sensitive to bias as parents rely on perceptions of infant sleep rather than actual infant sleep and may not always know when an infant is awake. Various factors could result in such reporting bias. For example, parenting stress, parent fatigue, and infant childcare attendance may all influence reporting [19, 20]. In contrast, objective methods including polysomnography and accelerometer are less prone to bias and have been increasingly used in infant populations. Polysomnography, typically measured in a laboratory setting, is thought to be the gold standard of sleep measurement [21]. Polysomnography determines sleep behaviors by a combination of measurements, including electroencephalography, electrooculography, electromyography, and electrocardiography [21]. An accelerometer, on the other hand, is a much simpler device that can be worn on the wrist, ankle, or waist and continuously records motility in the natural environment without the need for a laboratory. Many studies have used accelerometers to measure infant sleep, despite methodological challenges [9, 22, 23].

Multiple accelerometer devices are available and they have different placement requirements (e.g. ankle, wrist), settings of epoch length (e.g. 15, 60 seconds), and sensor sensitivity [22, 23]. This raises questions on device comparability of research findings [9, 22, 23]. Moreover, algorithms should be applied to raw accelerometer data before useful sleep parameters can be derived, but most of the available algorithms were not developed for use with infants. There are, however, four algorithms that have been applied in infant populations. The Sadeh algorithm was initially developed for use in young adults, with accelerometers attached to both wrists operating using the 60 seconds epoch [24]. When used in infants and compared to polysomnography, the Sadeh algorithm had a sensitivity (i.e. sleep agreement) of 89% and a specificity (i.e. wake agreement) of 52% with a single accelerometer placed on the ankle [25]. Sadeh et al. developed another algorithm specifically for infants with an accelerometer worn on the left ankle, referred to as the Sadeh Infant algorithm in this paper. This algorithm applied to accelerometer data in 60 seconds epoch had a sensitivity between 54.9% to 88.1% and a specificity between 82.8% to 99.3% relative to direct observation, depending on the specific age of infants [26]. Similarly, the Count-scaled algorithm was recently developed and validated against polysomnography in infants 10 to 22 weeks of age and had a sensitivity of 86% and specificity of 85% with accelerometer in the recommended 15 seconds epoch length [27]. The Count-scaled algorithm was argued to be independent of sensor sensitivities or placements [27]. The Cole algorithm, on the other hand, was developed in adults based on wrist actigraphy in 60 seconds epoch length [28]. It was reported to have a sensitivity of 80.0% and a specificity of 89.9% compared with polysomnography in infants [27].

Furthermore, in infant sleep research, researchers rely on specific definitions to derive sleep parameters after accelerometer data are scored using algorithms [22, 29]. However, existing definitions used by previous studies either are not reported or vary greatly [22, 29]. A review of 228 pediatric sleep studies with accelerometer assessment found that only 70% of studies provided definitions for the reported sleep parameters. The inconsistency of definitions was evident in the example of determining "bedtime/sleep onset" where studies have used 3, 10, 15, and 20 consecutive minutes of decreased activity as the definition [22]. Researchers have emphasized the critical importance of a comprehensive and clear reporting of the process of sleep parameter derivation based on accelerometer data to facilitate comparability [22, 29, 30]. Lastly, the aforementioned methodological complexities may have contributed to the inconsistent agreement between accelerometer and sleep diaries in measuring infant sleep reported by previous studies. Although some studies have reported good agreement [10, 17], others have reported poor or mixed agreement [15, 16, 18]. Studies have also documented that, compared to sleep diaries, accelerometers captured fewer daytime sleep and shorter duration of daytime sleep [9, 10, 15, 16, 30], while detecting greater frequency and longer duration of night wakings [9, 10, 14–16]. However, none of these studies evaluated the agreement while considering different algorithms, epoch lengths, and sleep parameter definitions used with accelerometer data simultaneously.

Our study aimed to address these research gaps by (1) assessing agreement of sleep/wake status identification by different algorithms based on accelerometer sampled at different epoch lengths, (2) deriving sleep parameters encompassing the 24-hour period using accelerometer based on recommended definitions, and (3) evaluating agreement of absolute values as well as longitudinal trajectories of sleep parameters obtained from sleep diary and accelerometer. Our hypotheses are: (1) the agreement of sleep/wake status identification based on accelerometer varies by algorithm and epoch length, (2) accelerometers tend to estimate fewer occurrences and shorter durations of daytime sleep but estimate higher frequency and longer duration of night waking than sleep diaries, and (3) accelerometer and sleep diaries estimate similar trajectories of sleep parameters. Although not originally developed for infants, a systematic review published in 2021 found that 10.1% and 4.1% of the 58 articles that assessed infant sleep via accelerometer used the Sadeh and Cole algorithm, respectively [23]. Therefore, we included these two algorithms in our study to facilitate a comprehensive comparison of different algorithms. We used data from a racially diverse cohort of infants residing in the southeastern United States with longitudinal measurements of sleep at 3, 6, 9, and 12 months.

Methods

Study design and population

Participants were from the Nurture study, a prospective, observational birth cohort of women and their infants residing in the southeastern United States (2013–2018) [31]. The Nurture study was designed to assess longitudinal associations of multiple caregivers on infant adiposity and weight trajectories throughout the first year of life. We recruited pregnant women between 20 to 36 weeks' gestation from a local private prenatal clinic and the county health department prenatal clinic. Women were eligible to participate if they had a singleton pregnancy with no known congenital abnormalities, were at least 18 years of age, spoke and read English, intended to keep the baby, and planned to stay within the area until at least 12 months postpartum. Infants were included if they were born after 28 weeks' gestation, did not have congenital abnormalities that could affect growth and development, were not in hospital for 3 or more weeks after birth, and were able to take food by mouth at the time of hospital discharge. We obtained written consent at recruitment during pregnancy and confirmed participation shortly after delivery for both mothers and infants. We conducted home visits when infants were aged 3, 6, 9, and 12 months. The Duke University Medical Center Institutional Review Board approved this study and its protocol. Additional information about the Nurture study is available elsewhere [31].

Sleep diaries

Mothers and other caregivers were instructed to concurrently complete a paper-based sleep diary for four full, continuous days over 2 weekdays and 2 weekend days, while infants were wearing the accelerometer at 3, 6, 9, and 12 months. The sleep diaries traveled with infants so the adult responsible was instructed to complete the diary (i.e. if the infant went to child care the child care provider completed the diary). Caregivers reported infant start and stop time for nighttime sleep and naps each day. We calculated sleep parameters including total nighttime sleep duration, frequency of wake after nighttime sleep onset (WASO) and naps, WASO duration, total WASO duration, nap duration, total daytime sleep, total nighttime sleep, nighttime sleep efficiency, and total 24-hour sleep, based on the definitions described in Table 1. In addition, caregivers noted the start and end times in sleep diaries when accelerometers were removed from infants.

Table 1. Decisions to Process Accelerometer Data and Derive Sleep Parameters in the Nurture Study

	Decisions
Non-Wear	Non-wear if: (1)recorded as non-wear in diary; or (2)had ≥60 consecutive minutes of 0 in accelerometer data between nighttime sleep offset and onset in a day.
Valid day	A 24-hour day was valid if had ≥12 hours of wear time.
Valid recording	A 4-day recording was valid if had ≥3 valid days.
Nighttime sleep onset	The start of the first ≥15 consecutive minutes of sleep in the time period between 09:00 pm to 09:00 am.
Nighttime sleep offset	The end of the last ≥15 consecutive minutes of sleep in the time period between 09:00 pm to 09:00 am.
Total nighttime sleep duration	Calculated as the duration of time between nighttime sleep onset and offset.
Wake after nighttime sleep onset (WASO)	In a WASO if had ≥5 consecutive minutes of wake between nighttime sleep onset and offset.
WASO duration	Calculated as the mean duration per WASO.
Total WASO duration	Calculated as the duration of all WASOs.
Nap	In a nap if had ≥15 consecutive minutes of sleep preceded by ≥5 consecutive minutes of wake between nighttime sleep offset and onset.
Nap duration	Calculated as the mean duration per nap.
Total daytime sleep	Time spent asleep between nighttime sleep offset and onset. Calculated as the duration of all naps.
Total nighttime sleep	Time spent asleep between nighttime sleep onset and offset. Calculated as (total nighttime sleep duration – total WASO duration).
Nighttime sleep efficiency	The percentage of time spent asleep between nighttime sleep onset and offset. Calculated as (total nighttime sleep ÷ total nighttime sleep duration ×100).
Total 24-hour sleep	Time spent asleep in a 24-hour day. Calculated as (total daytime sleep + total nighttime sleep).

Accelerometer

Infants wore ActiSleep+ accelerometer (ActiGraph, Pensacola, Florida, USA) on the left ankle for four full, continuous days at 3, 6, 9, and 12 months. We instructed caregivers to keep accelerometers on infants at all times including when infants were sleeping and bathing. The accelerometer used a 30 Hz sampling rate and recorded triaxial acceleration data in 15 seconds epoch.

Data cleaning and application of algorithms.

A total of 506 infants had accelerometer data available during at least one-time point, which in total comprised 5432 days. We downloaded the raw data using the ActiLife software (version 6.13.4) and converted it to activity counts in both 15 seconds and 60 seconds epochs. We classified non-wear time if the corresponding time period was labeled as non-wear in sleep diaries regardless of activity counts recorded by the accelerometer or if the accelerometer recorded ≥60 minutes of consecutive 0 seconds when not indicated as non-wear in sleep diaries, when the infants were not engaging in nighttime sleep (Table 1). We defined a valid day as having at least 12 hours of wear time and a valid recording as having at least 3 days of usable data [32]. We excluded 29 infants (comprising 438 days) due to insufficient valid days or valid recordings. Our final analytic sample included 477 infants (4994 days) who had at least one valid accelerometer recording across 4-time points. We then applied each of the four algorithms, including Sadeh [24], Sadeh Infant [26], Cole [28], and Count-scaled algorithm [27] separately for 15 and 60 seconds epochs.

Derivation of sleep parameters.

We derived a total of 12 sleep parameters based on the definitions described below and in Table 1. We calculated all sleep parameters for each infant for each day and then averaged across the days at 3, 6, 9, and 12 months. We determined *nighttime sleep onset* as the start of the first ≥15 consecutive minutes of sleep and *nighttime sleep offset* as the end of the last ≥15 consecutive minutes of sleep, between 09:00 pm to 09:00 am [18, 22]. We then calculated *total nighttime sleep duration* as the time elapsed between nighttime sleep onset and offset. We classified an infant as having a WASO if the accelerometer data had ≥5 consecutive minutes of wake time between nighttime sleep onset and offset [18, 22]. We then calculated *frequency of WASOs*, *WASO duration* as the mean duration per WASO, and *total WASO duration* as the duration of all WASOs. Next, we classified an infant as having a nap if the accelerometer data have ≥15 consecutive minutes of sleep preceded by ≥5 consecutive minutes of wake between nighttime sleep offset and onset [22, 30, 33]. We calculated *frequency of naps*, *nap duration*, and *total daytime sleep* following the similar approach as for WASO parameters. Lastly, we calculated *total nighttime sleep* as total nighttime sleep duration – total WASO duration, *nighttime sleep efficiency* as total nighttime sleep ÷ total nighttime sleep duration × 100%, and *total 24-hour sleep* as total daytime sleep + total nighttime sleep.

Other measures

We collected additional information on mothers and infants through interviews and questionnaires administered at recruitment, after delivery, and during each home visit. We obtained clinical information from a review of electronic medical records. Infant information included sex (male; female), race (black; white; other race or multiple races), ethnicity (not Hispanic or Latinx; Hispanic or Latinx), gestational age at birth (continuous), birth weight (continuous), delivery mode (not cesarean delivered; cesarean delivered), weeks of any breastfeeding from birth (continuous), and if shared bed with parents at the time of sleep measurement (yes; no). We further calculated infant birth weight for gestational age z-score based on the Intergrowth-21st standards [34]. Maternal information included age (continuous), race (black; white; other race or multiple races), ethnicity (not Hispanic or Latinx; Hispanic or Latinx), education (high school graduate or below; some college or above), marital status (married or living with partner; other), annual household income (<$20 000; ≥$20 000), parity (nulliparous; multiparous), and prepregnancy body mass index (continuous).

Statistical analyses

Agreement of sleep/wake status identification by different algorithms.

We examined agreement of sleep/wake status identification for each epoch between each pair of the algorithms, separately for accelerometer data in 15 and 60 seconds epoch during each home visit at 3, 6, 9, and 12 months. We calculated sleep agreement, wake agreement, and overall agreement for each infant in each day and then averaged these daily statistics across the recording for each infant at 3, 6, 9, and 12 months. Sleep (or wake) agreement was calculated as the proportion of epochs that were identified as sleep (or wake) by both algorithms among the epochs identified as sleep (or wake) by the second algorithm. Overall agreement was calculated as the proportion of epochs that were identified consistently by both algorithms among all epochs. We computed kappa and prevalence-adjusted, bias-adjusted kappa (PABAK) to assess level of agreement after accounting for the agreement expected by chance, as well as the unbalanced prevalence of sleep and wake in infants [35, 36]. The interpretation of kappa and PABAK is as follows: ≤0 indicates poor agreement, 0.01 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and 0.81 to 1 almost perfect agreement [35, 36].

Agreement of sleep parameters between sleep diaries and accelerometer.

We used mean and standard deviation (SD) to summarize all the 12 sleep parameters derived from sleep diaries and accelerometers in 15 and 60 seconds epochs based on 4 algorithms during each home visit at 3, 6, 9, and 12 months. We quantified agreement of each sleep parameter comparing accelerometer (based on a certain algorithm in either 15 or 60 seconds epoch) against sleep diaries using the Bland–Altman plot [37]. The Y axis of Bland–Altman plot represents the difference in the sleep parameter between accelerometer and sleep diaries with the middle dash line representing the mean difference; and the X axis represents the mean of the sleep parameter of these two measures. The upper and lower dash lines on Bland–Altman plots show the upper and lower limits of agreement within which 95% of the differences between the two measures are included. When the 95% confidence interval does not include 0, the sleep parameter estimated by accelerometer and sleep diaries are statistically significantly different.

Trajectories of sleep parameters over the course of infancy.

We built marginal regression models using the generalized estimation equation approach with robust variance and autoregressive 1 correlation structure to estimate monthly changes in sleep parameters over infancy. We used marginal linear regressions for the 10 sleep parameters that were continuous, including nighttime sleep offset and onset, nap and WASO duration, total nighttime sleep duration, total WASO duration, total daytime, nighttime, and 24-hour sleep, as well as nighttime sleep efficiency. We used marginal Poisson regressions for the remaining two sleep parameters that were count outcomes, including number of naps and WASOs. We estimated trajectories separately for sleep parameters derived from sleep diaries and accelerometer. We conducted all statistical analyses using R software version 4.0.2 (R Foundation for Statistical Computing). Evidence of statistical significance was based on a two-sided $p < .05$.

Results

We included a total of 477 infants in this study. Among the excluded infants who did not have sufficient accelerometer recordings (n = 29), we found that they were generally comparable to the included infants, except for being more likely to be Latinx, born via Cesarean section, born to a nulliparous mother, and having a shorter duration of any breastfeeding at 12 months (Supplementary Table S1). Infants who were excluded from the analysis due to attrition (n = 160) had a shorter duration of any breastfeeding at 12 months, were born to mothers who were younger and had lower prepregnancy body mass index; however, they were comparable to the included infants otherwise.

Among the included infants, 66.2% were black, 16.4% were white, 15.7% were other race or multiple races, and most (86.8%) were not Hispanic or Latinx (Table 2). A total of 50.5% of infants were male. On average, infants were born at 39.0 (SD: 1.6) weeks of gestation and weighed 3.2 (SD: 0.5) kg at birth. More than half (55.6%) of the infants were living in households with an annual income of <$20 000. In total, 390 infants had valid accelerometer data at 3 months, 334 at 6 months, 307 at 9 months, and 290 at 12 months. Among them, 355 infants also had available sleep diary data at 3 months, 291 at 6 months, 281 at 9 months, and 263 at 12 months. Over half (55.3%) of the infants shared beds with their parents at the time of sleep measurement. Mothers had a mean age of 27.7 (SD: 5.8) years and 54.1% of them attended some college or above.

Agreement of sleep/wake status identification by different algorithms

The algorithms applied to the 15 seconds epoch accelerometer data identified longer sleep duration and shorter wake duration than algorithms applied to the 60 seconds epoch data across all time points (Supplementary Table S2). The Sadeh Infant algorithm consistently identified the longest sleep duration regardless of epoch length, followed by the Cole, Sadeh, and the Count-scaled algorithms in the 15 seconds epoch data, or the Cole, Count-scaled, and Sadeh algorithms in the 60 seconds epoch data. Epoch-by-epoch comparisons of sleep/wake status identification between each pair of algorithms were similar across all time points (Figure 1). The highest sleep agreement was observed when comparing the Sadeh Infant algorithm against the other algorithms, as well as when comparing the Cole algorithm against the Sadeh and Count-scaled algorithms (Figure 1, first column). However, the Sadeh Infant algorithm had the lowest wake agreement with the other three algorithms (Figure 1, second column), while the wake agreement among the other three algorithms was similar. This led to similar overall agreements between the Sadeh, Cole, and Count-scaled algorithms, ranging from 85% to 95%. In contrast, the overall agreement was lower when comparing Sadeh Infant algorithm against all other three algorithms, ranging from 60% to 80% (Figure 1, third column). Similarly, the kappa and PABAK suggested that the Sadeh Infant algorithm had modest agreement with the other algorithms, while the other algorithms had moderate to almost perfect algorithms with each other (Figure 1, fourth and fifth columns).

Agreement of sleep parameters between sleep diaries and accelerometer

Accelerometers, regardless of epoch length and algorithm, tended to agree relatively well with sleep diaries with regard to nighttime sleep offset (Table 3, Supplementary Figure S1), nighttime sleep onset (Supplementary Figure S2), and total nighttime sleep duration (Supplementary Figure S3). However, there was a trend that the differences between accelerometer and sleep diaries were proportional to the size of nighttime sleep offset and onset in a

Table 2. Characteristics of Infants and Mothers in the Nurture Study (*n* = 477)

Infant Characteristics		
		Mean (SD)
Gestational age at birth, weeks		39.0 (1.6)
Birth weight, kg		3.2 (0.5)
Birth weight for gestational age z-score		0.1 (1.0)
Duration of any breastfeeding at 12 months, weeks		18.0 (18.3)
		N (%)
Female		236 (49.5%)
Race	Black	316 (66.2%)
	White	78 (16.4%)
	Other race or multiple races	75 (15.7%)
	Missing	8 (1.7%)
Ethnicity	Not Latinx	414 (86.8%)
	Latinx	41 (8.6%)
	Missing	22 (4.6%)
Cesarean delivered		152 (31.9%)
Co-slept with parents at the time of sleep measurement		264 (55.3%)
Maternal characteristics		
		Mean (SD)
Age, years		27.7 (5.8)
Prepregnancy BMI, kg/m^2		30.2 (9.1)
		N (%)
Race	Black	334 (70.0%)
	White	96 (20.1%)
	Other race or multiple races	46 (9.6%)
	Missing	1 (0.2%)
Ethnicity	Not Latina	449 (94.1%)
	Latina	27 (5.7%)
	Missing	1 (0.2%)
Education	High school or below	219 (45.9%)
	Some colleges or above	258 (54.1%)
Married or living with partner		275 (57.7%)
Annual household income	<$20 000	265 (55.6%)
	≥$20 000	191 (40.0%)
	Missing	21 (4.4%)
Parity	Nulliparous	163 (34.2%)
	Multiparous	300 (62.9%)
	Missing	14 (2.9%)

way that the differences were larger in infants who seemed to wake up or fall asleep either earlier or later than their peers.

As compared with sleep diaries, accelerometer using the 15 seconds epoch systematically estimated fewer number of naps (Table 3, Supplementary Figure S4 panel A). Using the 60 seconds epoch, the Sadeh algorithm generated similar number of naps to sleep diaries, the Sadeh Infant algorithm generated more naps, whereas the Cole and Count-scaled algorithms generated fewer naps (Supplementary Figure S4 panel B). Accelerometers systematically estimated shorter nap duration by about 70 minutes with 15 seconds epoch data and 50 minutes with 60 seconds epoch data (Table 3, Supplementary Figure S5). Consequently, accelerometers reported about 2 hours shorter total daytime sleep duration than sleep diaries, though the differences were smaller with the 60 seconds epoch based on the Sadeh or Sadeh Infant algorithms (Supplementary Figure S6). The discrepancies between accelerometer and sleep diaries were generally greater in infants who slept longer either per nap or in total throughout the day.

Accelerometers systematically estimated more number of WASOs but the magnitude was smaller when applying the Sadeh algorithm with the 60 seconds epoch or the Sadeh Infant algorithm with the 15 seconds epoch (Table 3, Supplementary Figure S7). The Cole and Count-scaled algorithms estimated more number of WASOs by about 3 counts with 15 seconds epoch data and 10 counts with 60 seconds epoch data (Supplementary Figure S7). Conversely, accelerometers derived on average about 20 minutes shorter duration per WASO than sleep diaries and the differences were smaller using an epoch length of 60 seconds instead of 15 seconds (Supplementary Figure S8). As for total WASO duration, it was similar to sleep diaries when using the 15 seconds epoch but longer when using the 60 seconds epoch based on algorithms other than Sadeh (Supplementary Figure S9). The differences in the three WASO-related sleep parameters between accelerometer and sleep diaries were also proportional to the sizes of these sleep parameters.

For total nighttime sleep and nighttime sleep efficiency, accelerometer, and sleep diaries agreed relatively well except when based on the Cole and Count-scaled algorithms using the 60 seconds epoch length, which estimated shorter total nighttime sleep by around 4 hours (Table 3, Supplementary Figure S10) and estimated higher nighttime sleep efficiency by around 25% (Supplementary Figure S11). The Sadeh Infant algorithm with the 60 seconds epoch slightly estimated higher nighttime sleep efficiency by roughly 10%. Finally, all algorithms estimated shorter total 24-hour sleep with the smallest difference being observed for the Sadeh Infant algorithm (Supplementary Figure S12). There were no clear patterns in total nighttime sleep or total 24-hour sleep comparing accelerometers with the 60 seconds epoch with sleep diaries.

Trajectories of sleep parameters over the course of infancy

Over the course of infancy, nighttime sleep offset remained statistically significantly unchanged and nighttime sleep onset tended to be 2.37 (95% CI: −3.46, −1.28) minutes earlier per month according to sleep diaries. However, these trajectories were shown to be mixed in accelerometers using different epoch lengths and algorithms (Figure 2, Supplementary Table S3). Infants had significantly fewer naps as they aged based on either sleep diaries or accelerometer, slept significantly longer in each nap based on sleep diaries and accelerometer using the 60 seconds epoch, and had significantly shorter total daytime sleep based on sleep diaries and accelerometer using the Sadeh Infant algorithm in either epoch length or the Sadeh and Cole algorithms using the 15 seconds epoch. For total daytime sleep, sleep diaries generated a decrement of 10.05 (95% CI: −11.39, −8.71) minutes per month, which was greater than the decrements generated by accelerometers (Figure 2, Supplementary Table S3).

As for WASO, accelerometers were largely consistent with sleep diaries in the direction and magnitude of change, indicating that infants had significantly fewer WASOs, shorter duration per WASO, as well as shorter total WASO duration from 3 months to 12 months. Sleep diaries and accelerometers in 60 seconds epoch were also in agreement with regard to significantly increasing total nighttime sleep and slightly increasing nighttime sleep efficiency over the course of infancy. Lastly, sleep diaries and the Sadeh Infant algorithm showed that infants had significantly lower total 24-hour sleep as they aged, whereas accelerometers based on other algorithms revealed either significantly longer or unchanged total 24-hour sleep (Figure 2, Supplementary Table S3).

Discussion

In 477 racially diverse infants measured longitudinally at 3, 6, 9, and 12 months, we found that epoch length and algorithm applied to accelerometer would impact the reliability of sleep/wake status identification. In general, the agreement between the Sadeh, Cole, and Count-scaled algorithms was higher than their agreement with the Sadeh Infant algorithm. As for sleep parameters relative to sleep diaries, accelerometer derived similar nighttime sleep offset, onset, and total nighttime sleep duration. Accelerometers also estimated similar total nighttime sleep and nighttime sleep efficiency when based on the Sadeh and Sadeh Infant algorithms, whereas accelerometer systematically estimated lower frequency and shorter duration of naps, estimated higher frequency of WASOs, and estimated shorter duration per WASO. Lastly, accelerometers and sleep diaries revealed some consistent trajectories of sleep parameters over infancy, including fewer naps and WASOs, longer duration per nap, shorter total daytime sleep, shorter duration per WASO, shorter total WASO duration, longer total nighttime sleep, and slightly increasing nighttime sleep efficiency. However, accelerometers and sleep diaries were not in agreement in the trajectories of nighttime sleep offset, onset, and total 24-hour sleep. The Sadeh Infant algorithm seemed to have more concordance with sleep diaries in terms of the estimation of sleep parameters, compared with the Sadeh, Cole, and Count-scaled algorithms.

A number of previous studies in the United States have investigated infant sleep using less objective methods, such as parent-reported sleep diary, interview, or questionnaire [9, 15]. However, most of these studies reported only a selected set of sleep parameters and did not capture the entire picture of infant sleep. Nevertheless, the sleep parameters derived in our study based on sleep diaries were largely comparable to the values reported by parents in previous studies in the United States [9, 15]. Furthermore, only a few studies in the U.S. employed a longitudinal design and examined changes in parent-reported infant sleep [30, 38]. McDaniel and Teti found in 150 infants from Pennsylvania that the number of WASOs from sleep diaries was greater when infants were aged 1 month compared to 3 months [38], which was in accordance with the trend observed in our study. In another study of 24 infants from Pennsylvania, Adams et al. reported similar trajectories to our findings that infants had fewer naps, shorter total daytime sleep, and longer total nighttime sleep from 6 to 24 weeks according to sleep diaries [30]. They observed no change in total 24-hour sleep—a trend that was also found by Sharkey et al. in 30 infants in Rhode Island measured at 2, 6, and 16 weeks [39]. Conversely, caregivers in our study reported a slight decrease in total 24-hour sleep from 3 to 12 months. Our study had a much larger sample size and more assessments than prior studies, which could explain the differences in findings. The decreasing total 24-hour sleep that we observed over infancy has also been reported in multiple studies in infants from outside of the United States based on parent reports [9].

The use of accelerometers in our study adds to a limited literature base of free-living objectively measured infant sleep in the United States [10, 14, 15, 30, 40]. Moreover, only one previous study had a somewhat comparable sample size [15, 40]. In 306 infants from Massachusetts measured at 1 and 6 months, the researchers derived six sleep parameters encompassing the 24-hour period based on accelerometer data [15, 40]. A few more sleep parameters also covering the full day were reported by Adams et al. in 24 infants from Pennsylvania [30]. However, several factors preclude us from directly comparing our results to these two prior studies. First, both of the studies adopted manual annotation conducted by researchers to score sleep/wake periods after applying the Cole [30] or Sadeh [40] algorithm in accelerometer data. We did not use manual scoring in our study, as this is not a feasible method for larger-scale longitudinal studies. Second, Adams et al. [30] measured infants at 6, 15, and 24 weeks, which only roughly aligned with the assessment time points of 3 and 6 months in our study. Though absolute values of sleep parameters differed between these studies [30, 40] and ours, some consistent trajectories were detected, including fewer naps, shorter total daytime sleep, fewer WASOs, and longer total nighttime sleep as infants aged.

Another three studies in the United States that used accelerometer reported only nighttime sleep parameters without including the daytime period. In 24 infants aged 12 months from Rhode Island, Acebo et al. applied the Sadeh Infant algorithm to 60 seconds epoch accelerometer data [10]. They reported an earlier mean nighttime sleep offset (6:52 for boys and 7:02 for girls vs. 8:01 overall in our study) but a later mean nighttime sleep onset (22:04 for boys and 22:51 for girls vs. 21:48 in our study), where we also used the Sadeh Infant algorithm with 60 seconds epoch accelerometer data for 12-month-old infants. One possible explanation for the discrepancies is the use of different definitions. Acebo et al. defined the nighttime period from 30 minutes before parents reported bedtime to 30 minutes after parents reported waking. Conversely, we defined the nighttime period as 21:00 to 9:00 without referring to caregiver report. In addition, they used a minimum of 3 and 5 consecutive minutes of sleep to define nighttime sleep onset and offset, respectively, whereas we used at least 15 consecutive minutes of sleep to define both. Furthermore, they found a slightly greater number of WASOs per night (7.6 for boys and 5.6 for girls vs. 5.88) using the definition of at least 3 consecutive minutes of awake time during nighttime sleep, compared to our study where we used at least 5 consecutive minutes. Horger et al. also used different definitions to define sleep parameters and their sample size was even smaller (nine infants) [14].

The existence of various definitions used by previous studies to derive sleep parameters based on accelerometer data adds to the complexity of comparing findings across studies. We followed most of the recommendations by Meltzer et al. noted in their review of 228 research articles that used accelerometer to measure sleep in pediatric population (41 in infants) [22]. We also adopted definitions from more recent evidence to define certain sleep parameters on which Meltzer et al. did not provide detailed recommendations. For example, a cutoff value of 20 consecutive minutes of sleep for identifying naps among infants aged 6 months has been reported to have better agreement with parent-reported sleep diaries than 30 or 40 minutes, but the agreement level was only slight [33]. Without further information from validation studies, we selected the value of 15 instead of 20 minutes in order to be more conservative in determining naps, also

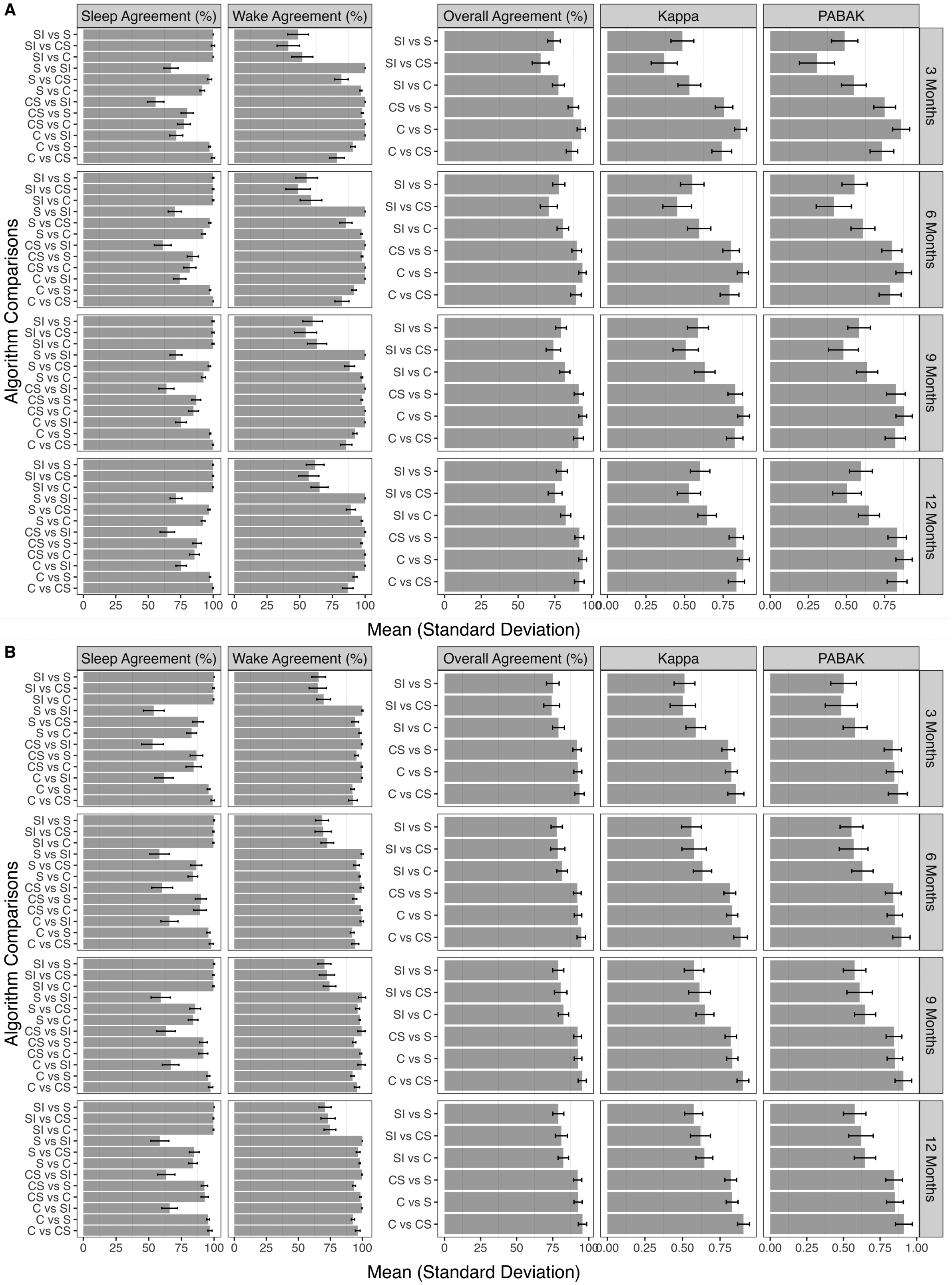


Figure 1. Comparison of sleep/wake status identification based on the Sadeh, Sadeh Infant, Cole, and Count-Scaled algorithms in accelerometer data at 15 seconds (A) and 60 seconds (B) epoch measured at 3, 6, 9, and 12 months of age in infants. Abbreviations: SI, Sadeh Infant algorithm; S, Sadeh algorithm; C, Cole algorithm; CS, Count-scaled algorithm; PABAK, prevalence-adjusted, bias-adjusted kappa. Sample size at 3 months is 390, 6 months is 334, 9 months is 307, and 12 months is 290.

Table 3. Sleep Parameters Derived From Sleep Diaries and Accelerometer in Infants Measured at 3, 6, 9, and 12 Months

	Nighttime Sleep Offset	**Nap *n*/day**	**Nap Duration minutes/nap**	**Total Daytime Sleep hours/day**	**Nighttime Sleep Onset**	**Total Nighttime Sleep Duration hours/day**	**WASO *n*/day**	**WASO Duration minutes/WASO**	**Total WASO Duration hours/day**	**Total Nighttime Sleep hours/day**	**Nighttime Sleep Efficiency %/day**	**Total 24-Hour Sleep hours/day**
Sleep diary												
3 months	7:53 (1:24)	2.85 (0.78)	85.46 (36.63)	3.89 (1.78)	22:12 (1:11)	9.59 (1.62)	1.28 (0.73)	28.61 (26.28)	0.81 (0.72)	8.83 (1.68)	91.69 (8.45)	12.82 (2.05)
6 months	7:54 (1:13)	2.36 (0.72)	79.41 (33.41)	2.98 (1.36)	22:00 (1:13)	9.85 (1.35)	0.95 (0.57)	25.50 (28.18)	0.57 (0.59)	9.36 (1.40)	94.26 (6.45)	12.24 (1.48)
9 months	8:05 (1:16)	1.91 (0.59)	84.19 (37.05)	2.55 (1.11)	21:52 (1:15)	10.10 (1.20)	0.92 (0.55)	19.62 (20.39)	0.45 (0.47)	9.60 (1.31)	95.53 (4.81)	12.19 (1.57)
12 months	8:05 (1:19)	1.61 (0.52)	96.51 (39.20)	2.46 (1.20)	21:55 (1:13)	10.06 (1.27)	0.72 (0.62)	17.86 (28.26)	0.37 (0.51)	9.72 (1.55)	96.02 (6.25)	12.08 (1.74)
Accelerometer, 15-second epoch												
Sadeh algorithm												
3 months	7:59 (0:40)	1.35 (0.67)	22.23 (11.20)	0.69 (0.42)	21:50 (0:29)	9.93 (0.85)	5.46 (2.00)	12.27 (4.63)	1.13 (0.52)	8.80 (0.87)	88.84 (4.97)	9.49 (0.97)
6 months	7:53 (0:44)	1.13 (0.63)	22.43 (14.30)	0.61 (0.41)	21:49 (0:33)	9.84 (0.90)	4.10 (1.75)	11.48 (5.06)	0.83 (0.46)	9.00 (0.86)	91.73 (4.40)	9.61 (1.02)
9 months	7:52 (0:46)	0.97 (0.54)	22.45 (18.20)	0.56 (0.40)	21:55 (0:36)	9.67 (0.77)	3.55 (1.58)	9.68 (4.31)	0.63 (0.36)	9.03 (0.75)	93.59 (3.53)	9.59 (0.88)
12 months	7:48 (0:47)	0.76 (0.52)	19.64 (15.35)	0.45 (0.36)	21:54 (0:36)	9.54 (0.83)	3.53 (1.53)	9.48 (4.97)	0.63 (0.41)	8.91 (0.90)	93.48 (4.22)	9.36 (1.01)
Sadeh infant algorithm												
3 months	8:32 (0:22)	1.00 (0.61)	12.27 (4.63)	0.64 (0.50)	21:31 (0:23)	10.95 (0.53)	2.19 (1.12)	7.57 (3.13)	0.34 (0.20)	10.61 (0.52)	96.94 (1.77)	11.25 (0.68)
6 months	8:19 (0:30)	1.05 (0.60)	11.48 (5.06)	0.73 (0.51)	21:34 (0:27)	10.67 (0.62)	1.75 (1.05)	6.74 (3.53)	0.27 (0.19)	10.40 (0.63)	97.48 (1.74)	11.12 (0.80)
9 months	8:16 (0:36)	0.87 (0.53)	9.68 (4.31)	0.63 (0.49)	21:41 (0:32)	10.43 (0.66)	1.27 (1.02)	5.21 (3.66)	0.20 (0.20)	10.22 (0.66)	98.10 (1.88)	10.85 (0.83)
12 months	8:15 (0:38)	0.75 (0.51)	9.48 (4.97)	0.54 (0.43)	21:41 (0:33)	10.36 (0.70)	1.22 (1.12)	4.79 (3.84)	0.20 (0.23)	10.16 (0.73)	98.13 (2.16)	10.70 (0.84)
Cole algorithm												
3 months	7:53 (0:43)	0.66 (0.49)	13.42 (11.18)	0.32 (0.29)	21:54 (0:29)	9.68 (0.91)	3.51 (1.51)	10.47 (3.65)	0.66 (0.36)	9.02 (0.89)	93.33 (3.47)	9.33 (0.96)
6 months	7:47 (0:46)	0.55 (0.43)	12.99 (11.51)	0.28 (0.26)	21:52 (0:34)	9.64 (0.98)	2.57 (1.36)	9.59 (4.68)	0.49 (0.33)	9.15 (0.93)	95.02 (3.22)	9.43 (1.01)
9 months	7:47 (0:46)	0.47 (0.41)	11.81 (12.72)	0.24 (0.26)	21:56 (0:35)	9.52 (0.83)	2.17 (1.21)	7.89 (4.11)	0.38 (0.28)	9.14 (0.80)	96.15 (2.81)	9.39 (0.88)
12 months	7:43 (0:48)	0.40 (0.41)	11.60 (12.94)	0.23 (0.27)	21:55 (0:36)	9.40 (0.83)	2.08 (1.18)	7.73 (4.69)	0.38 (0.34)	9.02 (0.89)	96.04 (3.63)	9.26 (0.98)
Count-scaled algorithm												
3 months	7:23 (0:55)	0.48 (0.42)	9.05 (7.54)	0.20 (0.20)	22:05 (0:34)	8.71 (1.24)	4.93 (1.92)	11.40 (4.87)	0.96 (0.47)	7.76 (1.16)	89.45 (4.81)	7.96 (1.21)
6 months	7:27 (0:51)	0.48 (0.41)	10.33 (9.56)	0.22 (0.21)	21:58 (0:35)	9.02 (1.10)	3.59 (1.55)	11.00 (4.98)	0.70 (0.41)	8.32 (1.01)	92.45 (4.10)	8.54 (1.08)
9 months	7:31 (0:53)	0.45 (0.44)	11.30 (15.39)	0.24 (0.31)	22:01 (0:37)	9.06 (0.93)	3.15 (1.42)	9.44 (4.86)	0.55 (0.32)	8.52 (0.90)	94.11 (3.40)	8.75 (0.99)
12 months	7:30 (0:52)	0.37 (0.36)	10.02 (11.66)	0.20 (0.23)	21:58 (0:37)	9.00 (0.98)	3.14 (1.43)	8.78 (4.67)	0.54 (0.34)	8.46 (1.03)	94.13 (3.69)	8.66 (1.10)
Accelerometer, 60-second epoch												
Sadeh algorithm												
3 months	7:25 (0:58)	2.56 (1.22)	21.73 (9.36)	1.21 (0.69)	22:03 (0:35)	8.87 (1.22)	1.60 (0.99)	11.36 (9.28)	0.42 (0.32)	8.45 (1.18)	95.31 (4.00)	9.67 (1.57)
6 months	7:25 (0:54)	2.49 (1.19)	23.97 (11.23)	1.25 (0.70)	21:58 (0:34)	9.01 (1.08)	1.86 (1.17)	8.75 (6.95)	0.38 (0.28)	8.63 (1.05)	95.91 (3.17)	9.88 (1.39)

Table 3. Continued

	Nighttime Sleep Offset	Nap *n*/day	Nap Duration minutes/ nap	Total Daytime Sleep hours/ day	Nighttime Sleep Onset	Total Nighttime Sleep Duration hours/day	WASO *n*/day	WASO Duration minutes/ WASO	Total WASO Duration hours/ day	Total Nighttime Sleep hours/ day	Nighttime Sleep Efficiency %/day	Total 24-Hour Sleep hours/day
9 months	7:27 (0:54)	2.23 (1.03)	24.89 (13.69)	1.19 (0.72)	22:04 (0:36)	8.94 (0.99)	1.84 (1.23)	8.59 (7.16)	0.38 (0.30)	8.57 (0.97)	95.89 (3.39)	9.75 (1.33)
12 months	7:27 (0:55)	2.07 (1.02)	26.29 (16.59)	1.15 (0.67)	22:00 (0:37)	8.92 (0.99)	1.93 (1.25)	8.13 (6.15)	0.38 (0.30)	8.54 (0.96)	95.83 (3.47)	9.69 (1.28)
Sadeh infant algorithm												
3 months	8:13 (0:34)	3.64 (1.04)	36.93 (11.90)	2.29 (0.76)	21:43 (0:27)	10.39 (0.71)	6.97 (2.37)	15.74 (5.20)	1.78 (0.71)	8.61 (0.89)	83.05 (6.64)	10.90 (1.19)
6 months	8:07 (0:38)	3.06 (0.92)	37.94 (13.87)	2.03 (0.78)	21:43 (0:31)	10.23 (0.78)	6.15 (2.35)	14.08 (6.05)	1.39 (0.62)	8.85 (0.85)	86.60 (5.79)	10.88 (1.16)
9 months	8:04 (0:43)	2.50 (0.93)	40.28 (19.33)	1.77 (0.72)	21:49 (0:35)	10.03 (0.74)	5.94 (2.23)	12.12 (6.18)	1.17 (0.52)	8.86 (0.78)	88.51 (5.02)	10.63 (1.07)
12 months	8:01 (0:44)	2.11 (0.81)	40.70 (18.38)	1.57 (0.68)	21:48 (0:35)	9.93 (0.76)	5.88 (2.09)	11.49 (5.65)	1.12 (0.51)	8.81 (0.93)	88.72 (5.17)	10.38 (1.12)
Cole algorithm												
3 months	7:31 (0:54)	1.97 (0.92)	20.35 (8.84)	0.89 (0.48)	22:02 (0:33)	9.04 (1.08)	9.21 (2.22)	19.65 (5.82)	2.93 (0.87)	6.11 (1.14)	67.79 (9.00)	6.99 (1.34)
6 months	7:29 (0:56)	1.92 (0.93)	22.47 (12.05)	0.94 (0.58)	21:57 (0:34)	9.12 (1.10)	10.24 (2.48)	15.39 (5.62)	2.53 (0.77)	6.59 (1.08)	72.28 (7.67)	7.53 (1.32)
9 months	7:32 (0:52)	1.72 (0.84)	23.73 (16.16)	0.89 (0.55)	22:03 (0:36)	9.08 (0.92)	11.22 (2.35)	13.73 (4.30)	2.52 (0.63)	6.55 (1.02)	72.03 (6.91)	7.44 (1.24)
12 months	7:29 (0:52)	1.58 (0.84)	23.92 (13.76)	0.86 (0.53)	21:58 (0:37)	8.97 (0.98)	11.36 (2.43)	14.57 (10.27)	2.54 (0.67)	6.43 (1.07)	71.58 (7.26)	7.28 (1.30)
Count-scaled algorithm												
3 months	7:18 (1:00)	1.99 (1.01)	18.20 (8.48)	0.85 (0.50)	22:05 (0:35)	8.59 (1.24)	10.15 (2.32)	20.96 (6.70)	3.43 (0.89)	5.17 (1.11)	60.67 (8.80)	6.02 (1.30)
6 months	7:25 (0:53)	2.07 (1.06)	22.17 (14.24)	0.99 (0.64)	22:00 (0:35)	8.96 (1.11)	11.64 (2.64)	15.77 (5.21)	2.96 (0.82)	6.01 (1.08)	67.13 (8.15)	7.00 (1.36)
9 months	7:28 (0:54)	1.95 (0.94)	23.15 (12.33)	0.99 (0.60)	22:04 (0:37)	9.00 (0.94)	12.15 (2.42)	14.04 (4.21)	2.79 (0.68)	6.21 (1.05)	68.83 (7.42)	7.20 (1.30)
12 months	7:28 (0:52)	1.78 (0.93)	23.69 (13.38)	0.95 (0.58)	21:58 (0:37)	8.92 (0.99)	12.33 (2.44)	14.47 (6.16)	2.80 (0.68)	6.12 (1.03)	68.39 (7.17)	7.06 (1.31)

Statistics are displayed using mean (standard deviation).
Sample size at 3 months is 355, 6 months is 291, 9 months is 281, and 12 months is 263.

taking into consideration that accelerometers are known to be more sensitive in detecting sleep than wake time [22]. Additionally, we included infants who wore accelerometers for at least 3 days, 12 hours per day (including daytime hours), to increase the validity of accelerometer data. This recommendation was made by a generalizability study conducted in 2017 on 23 infants between 1 to 12 months of age from Michigan [32]. In contrast, the previous five studies in the United States described above either failed to include daytime hours [10, 14] or only partially followed this recommendation [15, 30, 40].

Other challenges in comparing results across studies lies in inconsistent epoch lengths and algorithms used by previous research with accelerometer-based measurement of infant sleep. Our study provides valuable insights by evaluating four algorithms for identifying sleep/wake time with 15 and 60 seconds epochs. One study in 31 infants aged 10 to 22 weeks from New Zealand evaluated the performance of the Sadeh Infant, Cole, and Count-scaled algorithms using 15, 30, and 60 seconds epochs in a laboratory where infant sleep was measured during naps [27]. The researchers compared each algorithm against polysomnography and reported that all three algorithms generally yielded moderate agreement but the balance between sleep and wake agreement (i.e. sensitivity and specificity) was dependent on the specific algorithm and epoch length. Although that study did not compare each pair of algorithms, we found similar results in that the Sadeh Infant algorithm exhibited a higher sensitivity to sleep than wake, particularly when using 15 seconds epochs. This was evident through a higher agreement in sleep versus wake when compared to polysomnography, as well as when compared to other algorithms in our study. Furthermore, another study compared the Sadeh and Sadeh Infant algorithms using accelerometers with a 60 seconds epoch among nine infants aged approximately 1–8 months from the United States and Israel [14]. Consistent with our study, they found that the Sadeh Infant algorithm reported more sleep time than the Sadeh algorithm, even though they did not conduct epoch-by-epoch comparisons and thus did not report statistics such as sleep and wake agreement. Collectively, the higher agreement in sleep than wake by the Sadeh Infant algorithm compared to the Sadeh, Cole, and Count-scaled algorithms may be due to the Sadeh Infant

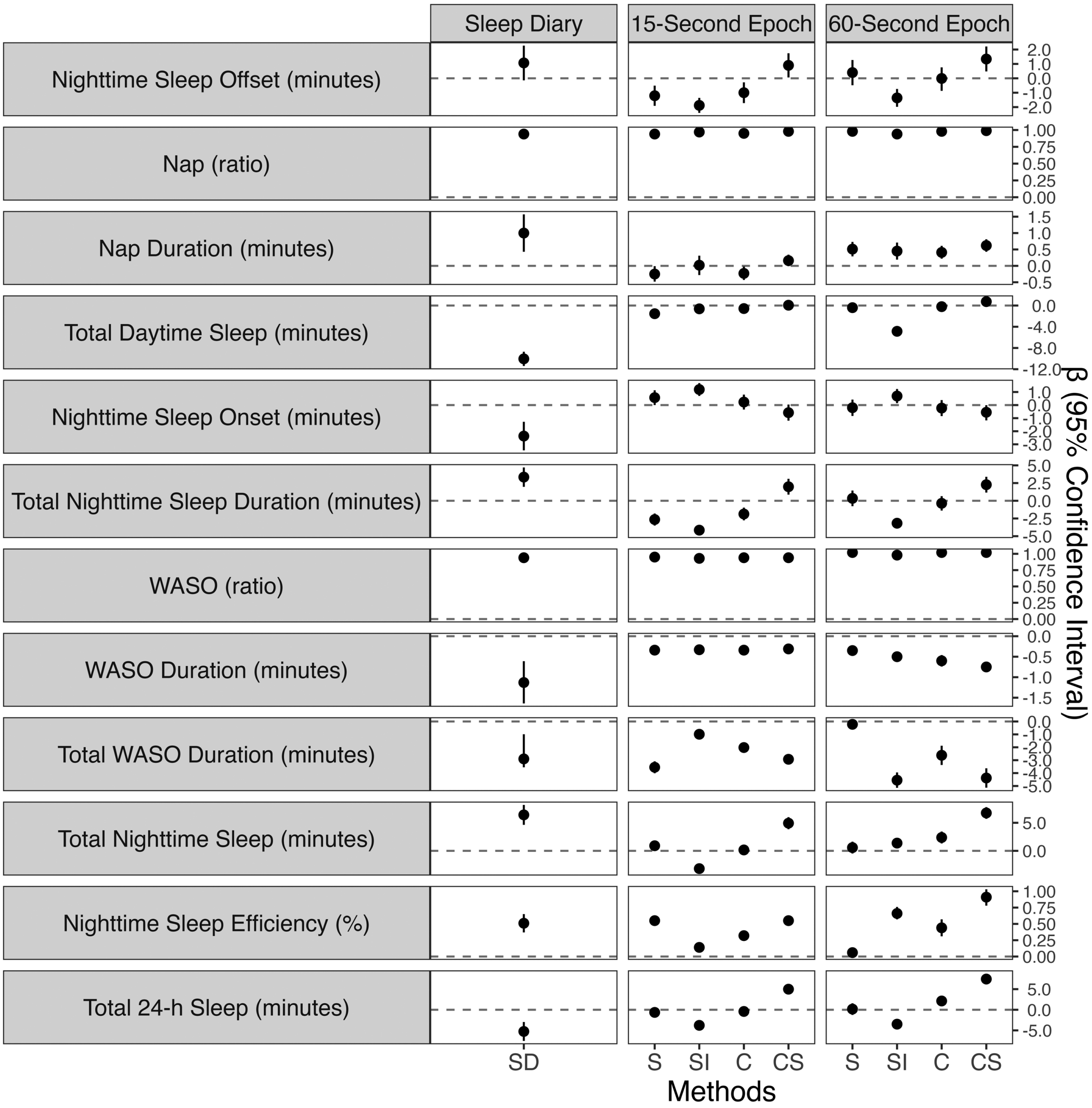


Figure 2. Estimated monthly changes in sleep parameters over the course of infancy, derived from sleep diaries, and accelerometer. Abbreviations: WASO, wake after sleep onset; SD, sleep diary; S, Sadeh algorithm; SI, Sadeh Infant algorithm; C, Cole algorithm; CS, Count-scaled algorithm. The estimates presented in this figure correspond to Supplementary Table S3.

algorithm's more stringent criteria for determining wake status based on infant movement. This aligns with the original intent behind developing the Sadeh Infant algorithm, which aimed to incorporate natural movement during infant sleep [26]. However, it is important to note that without a direct comparison to polysomnography or direct observation, our study cannot determine which algorithm demonstrates superior sleep/wake agreement.

The discrepancies we observed between caregiver-reported sleep diaries and accelerometers in measuring infant sleep were not unexpected given aforementioned methodological issues underlying accelerometers. Previous studies in and outside of the United States, along with ours, consistently suggested that relative to sleep diaries, accelerometers tended to underestimate frequency and duration of daytime sleep [9, 10, 15, 16, 30]. It should be noted that in our study, the underestimation by accelerometer was smaller in magnitude when using 60 seconds epoch versus 15 seconds epoch data. Nap duration appeared to drive the differences. One possible explanation could be that parents may not be able to instantly detect waking infants and thus would likely perceive slightly longer naps. The same reason could also help explain the overestimation in frequency and underestimation in duration of night wakening by accelerometer observed in our research and in previous studies [9, 10, 14–16]. Other possible factors that could influence caregiver perceptions of infant sleep include bed-sharing behaviors and mental health status of parents and especially mothers during postpartum period [19, 20].

Despite the fact that absolute values of sleep parameters differed to a varying extent between caregiver reports and accelerometers, the trajectories of sleep parameters over infancy had more congruence with studies in and outside of the United States [9]. We found that Sadeh Infant algorithm appeared to have more consistency with sleep diaries in terms of estimation of sleep parameters, as compared with the Sadeh, Cole, and Count-scaled algorithms. This could be due to the fact that the Sadeh Infant algorithm was initially developed and validated in infant populations, whereas the Sadeh and Cole algorithms were not.

Our study has limitations. First, although a previous study suggested that at least 5 nights were needed to obtain reliable measures of sleep in children [41], the Nurture study opted for a shorter period of wear time (4 continuous days) for two reasons. We recognized that infants would potentially struggle to wear the device for longer periods of time, as compared to studies of older children that recommended at least 5 days and we wanted to maximize adherence. Additionally, at the time of the study, we did not have specific guidance on wear time for infants related to sleep assessment so we estimated an appropriate length of time given challenges with assessment in infants. The second limitation of this study is that definitions used in our study to derive sleep parameters based on accelerometer data have not been fully validated in infants. However, we did our best to maximize the validity of our results by using definitions based on the most available evidence instead of from scratch. Third, we were not able to tease out acceleration generated by caregivers instead of infants which may have resulted in false wake time. This is inevitable in infant sleep research via accelerometer because infants are not fully mobile and no reliable approaches to date have been developed to automatically separate movements generated by infants and caregivers. Fourth, we did not compare parent-reported sleep diaries or accelerometers against polysomnography, as polysomnography was not used in the Nurture study. Therefore, our study should not be viewed as a validation study but rather a comparison between sleep diaries and accelerometers. A small study in the United States suggested that sleep diaries and accelerometers were more concordant with one another than with videosomnography (polysomnography with video) [13] and this should be investigated further in future studies. Fifth, even though our study is the largest in the United States to include objective longitudinal measures of infant sleep, infants in Nurture were not fully representative of infants in the United States, as there was a higher proportion of black race. Nevertheless, our study makes a valuable contribution because black participants are underrepresented in most U.S. birth cohorts [42], and in the only other larger study of infant accelerometers [40], only 8% of the 306 infants were black. Sixth, we did not evaluate differences in terms of infants who regularly did and did not co-sleep with their parents. Finally, as with most birth cohorts in which families experienced rapid change due to the birth of an infant, we faced issues with attrition. However, from birth to 12 months, we retained over 70% of the Nurture sample. This retention rate is higher than that of a similar birth cohort from the same city, in which only 56% of women completed the 12-month visit [43].

Conclusion

Our study provides the first comprehensive investigation in the United States on both reference values and longitudinal changes in 24-hour sleep in infants using subjective and more objective methods. We found that the reliability of sleep/wake identification was dependent on epoch length of accelerometers and algorithms used. These two factors also influenced the agreement of accelerometers with caregiver-reported sleep diaries in absolute values, as well as monthly changes in sleep parameters. However, researchers should use caution when interpreting these results, as both sleep diaries and accelerometers have limitations when used to measure infant sleep. The use of the Sadeh and Cole algorithms in infants should also be approached with caution, as they were originally developed for use in adults. Future efforts should be made to continue developing, improving, and validating algorithms and definitions that can be applied to accelerometer data to more reliably estimate infant sleep over the first year of life. Our findings support guidance put forth in previous literature reviews [22, 23], including recommendations related to device, epoch length, algorithm, definitions of sleep parameters, and data processing procedures. Future research on infant sleep using accelerometers should consider these recommendations to improve validity and enhance comparability.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Acknowledgments

We thank the caregivers and infants who participated in the Nurture study. We also thank Dr. Jiawei Bai (Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health) for his assistance in accelerometer data processing.

Funding

This study was supported by a grant from the National Institutes of Health (R01DK094841). The funders had no role in the design of the study, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure Statement

None declared.

Data Availability Statement

The data that support the findings of this study are available from Dr. Benjamin-Neelon, the Nurture birth cohort principal investigator. Some restrictions apply to the availability of these data. Data may be made available upon reasonable request and with permission from Duke University Medical Center.

References

1. Rudzik AEF, Ball HL. Biologically normal sleep in the mother-infant dyad. *Am J Hum Biol.* 2021;**33**(5):e23589.
2. Mindell JA, Sadeh A, Wiegand B, How TH, Goh DY. Cross-cultural differences in infant and toddler sleep. *Sleep Med.* 2010;**11**(3):274–280. doi: 10.1016/j.sleep.2009.04.012
3. Barry ES. What is "normal" infant sleep? why we still do not know. *Psychol Rep.* 2021;**124**(2):651–692. doi: 10.1177/0033294120909447
4. Zuckerman B, Stevenson J, Bailey V. Sleep problems in early childhood: continuities, predictive factors, and behavioral correlates. *Pediatrics.* 1987;**80**(5):664–671.
5. Byars KC, Yolton K, Rausch J, Lanphear B, Beebe DW. Prevalence, patterns, and persistence of sleep problems in the first 3 years of life. *Pediatrics.* 2012;**129**(2):e276–e284. doi: 10.1542/peds.2011-0372
6. Lam P, Hiscock H, Wake M. Outcomes of infant sleep problems: a longitudinal study of sleep, behavior, and maternal well-being. *Pediatrics.* 2003;**111**(3):e203–e207. doi: 10.1542/peds.111.3.e203
7. Gregory AM, Caspi A, Eley TC, Moffitt TE, O'Connor TG, Poulton R. Prospective longitudinal associations between persistent sleep problems in childhood and anxiety and depression disorders

in adulthood. *J Abnorm Child Psychol.* 2005;**33**(2):157–163. doi: 10.1007/s10802-005-1824-0

8. Cook F, Conway LJ, Giallo R, Gartland D, Sciberras E, Brown S. Infant sleep and child mental health: a longitudinal investigation. *Arch Dis Child.* 2020;**105**(7):655–660. doi: 10.1136/archdischild-2019-318014
9. Dias CC, Figueiredo B, Rocha M, Field T. Reference values and changes in infant sleep-wake behaviour during the first 12 months of life: a systematic review. *J Sleep Res.* 2018;**27**(5):e12654. doi: 10.1111/jsr.12654
10. 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**(12):1568–1577. doi: 10.1093/sleep/28.12.1568
11. Asaka Y, Takada S. Comparing sleep measures of infants derived from parental reports in sleep diaries and acceleration sensors. *Acta Paediatr.* 2011;**100**(8):1158–1163. doi: 10.1111/j.1651-2227.2011.02204.x
12. Simard V, Bernier A, Bélanger M-E, Carrier J. Infant attachment and toddlers' sleep assessed by maternal reports and actigraphy: different measurement methods yield different relations. *J Pediatr Psychol.* 2013;**38**(5):473–483.
13. Camerota M, Tully KP, Grimes M, Gueron-Sela N, Propper CB. Assessment of infant sleep: how well do multiple methods compare? *Sleep.* 2018;**41**(10).
14. Horger MN, Marsilliani R, DeMasi A, Allia A, Berger SE. Researcher choices for infant sleep assessment: parent report, actigraphy, and a novel video system. *J Genet Psychol.* 2021;**182**(4):218–235..
15. Quante M, Hong B, von Ash T, *et al.* Associations between parent-reported and objectively measured sleep duration and timing in infants at age 6 months. *Sleep.* 2021;**44**(4). doi: 10.1093/sleep/zsaa217
16. Gossé L, Wiesemann F, Elwell C, Jones E. Concordance between subjective and objective measures of infant sleep varies by age and maternal mood: Implications for studies of sleep and cognitive development. *Infant Behav Dev.* 2022;**66**:101663. doi: 10.1016/j.infbeh.2021.101663
17. Müller S, Hemmi MH, Wilhelm FH, Barr RG, Schneider S. Parental report of infant sleep behavior by electronic versus paper-and-pencil diaries, and their relationship to actigraphic sleep measurement. *J Sleep Res.* 2011;**20**(4):598–605. doi: 10.1111/j.1365-2869.2011.00926.x
18. Tikotzky L, Volkovich E. Infant nocturnal wakefulness: a longitudinal study comparing three sleep assessment methods. *Sleep.* 2018;**42**(1). doi: 10.1093/sleep/zsy191
19. Gress-Smith JL, Luecken LJ, Lemery-Chalfant K, Howe R. Postpartum depression prevalence and impact on infant health, weight, and sleep in low-income and ethnic minority women and infants. *Matern Child Health J.* 2012;**16**(4):887–893. doi: 10.1007/s10995-011-0812-y
20. Cook F, Conway L, Gartland D, Giallo R, Keys E, Brown S. Profiles and predictors of infant sleep problems across the first year. *J Dev Behav Pediatr.* 2020;**41**(2):104–116. doi: 10.1097/DBP.0000000000000733
21. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Med Rev.* 2020;**49**:101227. doi: 10.1016/j.smrv.2019.101227
22. Meltzer LJ, Montgomery-Downs HE, Insana SP, Walsh CM. Use of actigraphy for assessment in pediatric sleep research. *Sleep Med Rev.* 2012;**16**(5):463–475. doi: 10.1016/j.smrv.2011.10.002
23. Schoch SF, Kurth S, Werner H. Actigraphy in sleep research with infants and young children: current practices and future benefits of standardized reporting. *J Sleep Res.* 2021;**30**(3):e13134. doi: 10.1111/jsr.13134
24. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep.* 1994;**17**(3):201–207. doi: 10.1093/sleep/17.3.201
25. Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *J Sleep Res.* 2009;**18**(1):85–98. doi: 10.1111/j.1365-2869.2008.00706.x
26. 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 Dev.* 1995;**18**(3):329–337. doi: 10.1016/0163-6383(95)90021-7
27. Galland BC, Kennedy GJ, Mitchell EA, Taylor BJ. Algorithms for using an activity-based accelerometer for identification of infant sleep-wake states during nap studies. *Sleep Med.* 2012;**13**(6):743–751. doi: 10.1016/j.sleep.2012.01.018
28. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep.* 1992;**15**(5):461–469. doi: 10.1093/sleep/15.5.461
29. Galland B, Meredith-Jones K, Terrill P, Taylor R. Challenges and emerging technologies within the field of pediatric actigraphy. *Front Psychiatry.* 2014;**5**:99. doi: 10.3389/fpsyt.2014.00099
30. Adams EL, Master L, Buxton OM, Savage JS. A longitudinal study of sleep-wake patterns during early infancy using proposed scoring guidelines for actigraphy. *Sleep Med.* 2019;**63**:98–105. doi: 10.1016/j.sleep.2019.05.017
31. Benjamin Neelon SE, Ostbye T, Bennett GG, *et al.* Cohort profile for the Nurture Observational Study examining associations of multiple caregivers on infant growth in the Southeastern USA. *BMJ Open.* 2017;**7**(2):e013939. doi: 10.1136/bmjopen-2016-013939
32. Pitchford EA, Ketcheson LR, Kwon HJ, Ulrich DA. Minimum accelerometer wear time in infants: a generalizability study. *J Phys Act Health.* 2017;**14**(6):421–428. doi: 10.1123/jpah.2016-0395
33. Galland B, Meredith-Jones K, Gray A, *et al.* Criteria for nap identification in infants and young children using 24-h actigraphy and agreement with parental diary. *Sleep Med.* 2016;**19**:85–92. doi: 10.1016/j.sleep.2015.10.013
34. Villar J, Cheikh Ismail L, Victora CG, *et al.*; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet.* 2014;**384**(9946):857–868. doi: 10.1016/S0140-6736(14)60932-6
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;**33**(1):159–174.
36. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;**85**(3):257–268.
37. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;**327**(8476):307–310.
38. McDaniel BT, Teti DM. Coparenting quality during the first three months after birth: the role of infant sleep quality. *J Fam Psychol.* 2012;**26**(6):886–895. doi: 10.1037/a0030707
39. Sharkey KM, Iko IN, Machan JT, Thompson-Westra J, Pearlstein TB. Infant sleep and feeding patterns are associated with maternal sleep, stress, and depressed mood in women with a history of major depressive disorder (MDD). *Arch Womens Ment Health.* 2016;**19**(2):209–218. doi: 10.1007/s00737-015-0557-5
40. Yu X, Quante M, Rueschman M, *et al.* Emergence of racial/ethnic and socioeconomic differences in objectively measured

sleep-wake patterns in early infancy: results of the Rise & SHINE study. *Sleep.* 2021;**44**(3).

41. Acebo C, Sadeh A, Seifer R, *et al*. Estimating sleep patterns with activity monitoring in children and adolescents: how many nights are necessary for reliable measures? *Sleep.* 1999;**22**(1):95–103. doi: 10.1093/sleep/22.1.95
42. Konkel L. Racial and ethnic disparities in research studies: the challenge of creating more diverse cohorts. *Environ Health Perspect.* 2015; **123**(12): A297–A302.
43. Palmu J, Salosensaari A, Havulinna AS, *et al*. Association between the gut microbiota and blood pressure in a population cohort of 6953 individuals. *J Am Heart Assoc.* 2020;**9**(15):e016641.