SLEEPJ, 2019, 1–12

doi: 10.1093/sleep/zsy191
Advance Access publication Date: 4 October 2018
Original Article

Original Article

Infant nocturnal wakefulness: a longitudinal study comparing three sleep assessment methods

Liat Tikotzky* and Ella Volkovich

Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Work Performed: Department of Psychology, Ben-Gurion University of the Negev, Israel.

*Corresponding author. Liat Tikotzky, Department of Psychology, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 8410501, Israel. Email: liatti@bgu.ac.il.

Abstract

Study Objectives: To examine the convergence between actigraphy, sleep diaries, and the Brief Infant Sleep Questionnaire (BISQ) in the assessment of infant nocturnal wakefulness (i.e. minutes awake after sleep onset [WASO] and number of night-wakings [NW]) in the context of a longitudinal study.

Methods: The sample included 226 families, who were recruited during pregnancy. Data were collected at 3, 6, 12, and 18 months postpartum. Infants' sleep was monitored at home for five nights using actigraphy, sleep diaries, and the BISQ. Outcome measures included WASO and NW, as well as sleep latency, sleep duration, and sleep onset.

Results: Trajectory analyses demonstrated that all three methods showed declines in NW and WASO from 3 to 18 months. Statistically significant correlations were found between the three methods at all assessment points for all sleep variables. However, agreement rates (using Krippendorff's α and Bland–Altman analyses) between actigraphy and parental reports were poor. For NW, agreement between actigraphy and parental reports at 18 months was lower than that at 3 and 6 months. Diaries and BISQ showed satisfactory agreement for sleep latency.

Conclusions: Although the three methods' measures of infant nocturnal wakefulness are significantly correlated during infancy, absolute agreement between these methods is poor overall. The growing disagreement between actigraphy and parental reports (in NW) across development probably suggests that parents become less aware of infants' awakenings, due to the increasing ability of infants to self-soothe. Using both objective and subjective assessment methods seems especially important after the age of 6 months.

Statement of Significance

The reliable evaluation of infants' night-waking problems is crucial for clinical and research purposes. This longitudinal study demonstrated, for the first time, how actigraphy, sleep diaries, and the Brief Infant Sleep Questionnaire converge in the assessment of infant nocturnal wakefulness from 3 to 18 months. Examining the agreement between different sleep assessment methods in a longitudinal context is necessary, because the convergence may change along infant development. In the present study, all three methods were significantly correlated and clearly demonstrated that infant sleep consolidated over time. However, absolute agreement between the three methods was poor overall, and worsened over time for the number of awakenings. Future studies should examine the correspondence between different sleep assessment methods in clinically sleep-disturbed infants.

Key words: infant sleep; nocturnal wakefulness; actigraphy; sleep diaries; BISQ; longitudinal; trajectory

Submitted: 29 May, 2018; **Revised:** 27 August, 2018

Introduction

The consolidation of sleep–wake patterns is a major developmental task in infancy, influenced by complex interactions between physiological and socioenvironmental processes [1, 2]. Although most infants achieve the goal of uninterrupted nighttime sleep during the first year of life, infant sleep problems, manifested mainly in the form of frequent and/or prolonged night-wakings (NWs), characterize as many as 20%–30% of infants and toddlers, and are among the most prevalent complaints that parents present to pediatric health care professionals [3, 4].

Given the high prevalence of NW problems, their persistence and potential negative implications on child and family functioning [5–8], the reliable identification of these problems is crucial for both clinical and research purposes. Different methods are used to assess sleep in infants, including objective methods such as polysomnography, actigraphy, videosomnography, and self-reported measures, such as sleep diaries and questionnaires [9–12]. The present article focuses on three commonly used sleep assessment methods in early childhood, namely questionnaires, sleep diaries, and actigraphy. Actigraphy is a useful method that provides objective, noninvasive, continuous assessment of activity-based, sleep–wake patterns in one's natural sleep environment [11, 13, 14]. The validity of actigraphy against polysomnography and direct sleep observations has been demonstrated in various studies in the pediatric population [15–18]. However, actigraphy has been criticized for its relatively low specificity (i.e. wake detection ability) with some devices and some scoring algorithms [13, 18–20]. The advantages of parental reports of infant sleep by questionnaires or sleep diaries are clear: They are easy to administer, cheap, and provide data about bedtime behaviors that cannot be acquired by actigraphy (e.g. parental soothing behaviors, such as rocking, feeding) [9, 21]. In comparison to questionnaires, sleep diaries, which are based on day-to-day descriptions, are less likely to be biased by recall. However, parents may find it difficult to persist with their completion [10, 12, 22]. Parental subjective reports seem to also be limited by the restricted knowledge parents may have about their child's nocturnal awakenings. Parents are usually aware of NWs that involve infant signaling (e.g. crying). However, by the age of 6 months, most infants develop the ability to resume sleep after awakening at night without calling for their parents' attention [15, 23–25].

The question of the degree of convergence between these different sleep assessment methods has been addressed in several former studies. Overall, studies comparing actigraphy and subjective sleep measures agree that the correspondence between these methods is high for sleep schedule measures (e.g. sleep onset), but relatively low when sleep quality measures (e.g. number and length of NWs) are considered [11, 22, 23, 26]. Generally, it seems that parental estimation of infant nighttime wakefulness is shorter in comparison to actigraphy [10, 20, 22, 27, 28]. However, it is unclear whether these discrepancies are mainly a result of incorrect parental reports, of parents being unaware of infants' quiet awakenings, or are related to the inaccurate detection of wakefulness by actigaphy [13, 20].

Most studies comparing the agreement of different sleep assessment methods have been cross-sectional and only a few studies have examined the level of agreement along different stages of infant development [27, 29]. To our knowledge, there are no longitudinal studies comparing in parallel the convergence of questionnaires, prospective sleep diaries, and actigraphy. Examining the discrepancies and agreement between different sleep assessment methods in a longitudinal context is important, because the convergence between these methods may change as a factor of infant maturation and development of sleep-related behaviors. For instance, it could be that the disagreement between actigraphy and parental reported measures would increase with age when infants develop self-soothing capacities, reflected in the growing ability to independently resume sleep after nocturnal awakenings [25].

Previous studies of infant sleep development have demonstrated a clear pattern of infant sleep consolidation over the first year of life [1, 2, 27, 29–32]. However, these studies have relied mostly on one method to assess changes in sleep consolidation (usually parental reports). Thus, it is unclear whether different assessment methods would reveal a similar pattern of change across development.

Accordingly, the current study had two main aims. The first aim was to examine how well do actigraphy, sleep diaries, and a sleep questionnaire agree in the assessment of infant nocturnal wakefulness at different time points (i.e. 3, 6, 12, and 18 months) in the context of a longitudinal study. To that end, we first examined both mean differences between the three methods and their correlations at each assessment point. We also compared the strengths of the correlations across time points, expecting to find higher associations between actigraphy and parents at younger ages than older ages. Because correlations and mean values are only part of the information required to decide about agreement between methods, we further examined the rate of agreement between our methods with the statistical approach for assessing agreement between two methods of clinical measurement, developed by Bland and Altman [12, 33].

The second aim of this study was to examine and compare the developmental trajectories of infant nighttime wakefulness from 3 to 18 months using the three different assessment methods. We hypothesized that all three methods would reveal an increase in infant sleep consolidation over time. A unique aspect of this study was the comparison of the three methods' trajectories, allowing us to examine whether the trajectories would differ in their growth components.

Although we were mainly interested in investigating infant nocturnal wakefulness and sleep consolidation, we took advantage of our data to examine the level of concordance and developmental change for a few additional infant sleep measures: sleep latency, sleep duration, and sleep onset.

Methods

Participants

The present study included 226 families, who were recruited during pregnancy through prenatal courses and announcements on internet forums. The sample was enrolled during pregnancy to avoid recruitment biases related to infants' sleep difficulties. Only two parent families expecting their first child with a singleton full-term pregnancy and a healthy infant participated in the study. We decided to focus on families with a single child to reduce variance related to having prior experience with parenting that could influence the infants' sleep–wake patterns.

Table 1. Sample characteristics

Demographic	Mean ± SD	%
Mother's age (years)	28.81 ± 3.28	
Mother's education (years)	15.65 ± 2.02	
Father's age (years)	31.66 ± 7.04	
Father's education (years)	14.99 ± 2.32	
Maternal delivery week	39.26 ± 1.3	
Number of rooms at home	3.32 ± 0.92	
Infant weight 3 months (kg)	5.88 ± 0.80	
Infant weight 6 months (kg)	7.70 ± 0.95	
Infant weight 12 months (kg)	9.45 ± 1.2	
Infant weight 18 months (kg)	11.03 ± 1.1	
Child sex		50.5% boys

Demographic characteristics (parental age and education) were collected during the third trimester of pregnancy (Table 1). The study included four assessment points at 3, 6, 12, and 18 months postpartum. The *N* at 3 months was 226, at 6 months, *N* was 191, at 12 months, *N* was 172, and at 18 months, *N* was 150. Main reasons for discontinuation were lack of willingness to participate, overload (e.g. were too busy at work and/or at home), moving abroad and health problems. In addition, we had to discontinue the participation of 7 families at 12 months and of 16 families at 18 months because the study had ended (in terms of funding). The families who withdrew from the study were compared to the participating families on sociodemographic variables. No differences were found in any of these variables.

Procedures

The study was approved by the Helsinki committee of Soroka Medical Center in Israel. All parents signed informed consent before the first assessment. At each assessment point (3, 6, 12, and 18 months), a research assistant arrived at the participants' home and instructed them in actigraphy use. We chose the age of 3 months as our starting point because, at this age, day–night circadian rhythms are already quite organized [34–36] allowing most parents to clearly define the beginning of the nocturnal sleep period. This was important because our study focused on nighttime sleep and not on the 24-h sleep–wake cycles. Infant sleep was assessed for five nights (excluding weekends), using actigraphy and sleep diaries. In addition, parents completed the Brief Infant Sleep Questionnaire (BISQ) during the sleep assessment week. After completing the assessments, participants received a small gift (value of about $20) and a graphic report of their infant's actigraphic sleep.

Measures

Actigraphy

The actigraph registers motility data, which is translated into sleep–wake measures based on a computerized scoring algorithm [15, 37]. In the present study, we used the micromotion logger sleep watch (Ambulatory Monitoring, Inc., Ardsley, NY) with a 1-min epoch interval according to the standard working mode for sleep–wake scoring. The Actigraphic Sleep Analysis program was used to score the data based on Sadeh's validated scoring algorithm for infants [16]. The measures included in the present study were: (1) NW—number of night-wakings that last for at least 5 min; (2) wake after sleep onset (WASO)—minutes awake during the night; (3) sleep onset—defined by the first minute of the first consecutive 15 min of sleep after bedtime; and (4) sleep duration—hours in bed from sleep onset time to morning awakening time, including nocturnal wakefulness. All measures were averaged across the monitoring period. Most infants had valid actigraphy data for the full five assessment nights in all study phases. However, because of technical problems (e.g. actigraph fell off in the middle of the night), some data were missing. Thus, at 3 months, we had five nights of actigraphy data for 70.3% of the infants; four nights for 21.2%; and three nights for 8.5%. At 6 months, the rates were: 5 nights—63.0%; 4 nights—23.8%; and 3 nights—13.2%. At 12 months, the rates were: 5 nights—75%; 4 nights—22.7%; and 3 nights—2.3%. Last, at 18 months, the rates were: 5 nights—67.3%; 4 nights—20.4%; and 3 nights—12.3%.

Sleep diaries

Parents completed a nightly report on their infant's sleep patterns in parallel to the actigraphic assessment [22]. Parents were instructed to complete the sleep diary in real-time (i.e. mark it during the night when an awakening happened). The measures included in the present study were: (1) NW—number of night-wakings of any length; (2) WASO—minutes awake during the night; (3) sleep onset time; (4) sleep duration—the interval between parental reported sleep onset time and morning awakening time; and (5) sleep latency—time in minutes it took the infant to fall asleep. All measures were averaged across the assessment period.

Brief Infant Sleep Questionnaire (BISQ)

The BISQ is a well-validated questionnaire aimed at assessing infants' averaged sleep patterns [10]. Parents are instructed to refer to the infant's sleep patterns in the present. The derived measures for the purpose of the current study were: (1) NW—number of night-wakings of any length; (2) WASO—minutes awake during the night; (3) sleep onset time; and (4) sleep latency. Sleep duration could not be calculated because the BISQ does not include any questions on morning awakening time.

Background questionnaires

Sociodemographic and developmental data were collected, including parental age and education, number of rooms at home, delivery week, infant gender, and infant weight.

Analysis plan

We conducted separate, parallel analyses for each infant sleep measure. The first stage of each analysis was to examine concordance among the three (or two, as applicable) methods for each sleep variable at each assessment point. Correlations reflecting relative consistency and ANOVA models reflecting absolute differences were estimated in *Mplus* v8.1 [38], using that software's facility for accommodating missing data. Bland–Altman plots for absolute concordance for continuous measures [33, 39] were generated in R v3.5.0 [40], using the package *MethComp* v.1.22.2 [41]. Krippendorff's α for absolute concordance among ordinal variables were assessed using the R package *rel* v.1.3.1 [42].

Longitudinal analyses were conducted in *Mplus*. All *Mplus* analyses used the "MLR" estimator. MLR is a maximum likelihood estimator that is reasonably robust to deviations from multivariate normality. For each sleep variable, we estimated latent trajectory models (LTM) for the multiple methods (actigraphy, diary, and BISQ) in a single model. We built up each model stepwise from the simplest (latent trait) to the most complex (latent basis) available LTMs [43] as needed for model fit. We then contrasted the trajectory parameters (intercept, linear slope, etc.) across the three methods. Pairwise differences were adjusted to constrain the false discovery rate (FDR) with the Benjamini–Hochberg procedure [44].

Missing data and data preparation

MLR in *Mplus* uses all cases to maximize statistical power and is unbiased when item-level missing data are missing at random [45]. Bland–Altman plots apply listwise deletion. The plots also do not easily accommodate bivariate outliers. Before creating the Bland–Altman plots, we used box plots to identify univariate outliers. Testing for multivariate outliers was not feasible with the relatively low ratio of respondents to analysis variables. We converted outlying values on each continuous sleep measure (i.e. excluding NW, see below) to missing based on box plots. For other continuous-data analyses, we retained all observations, using the percentile bootstrap with 3,000 draws to generate estimates and statistical tests that accommodate the non-normal distributions.

NW is a count variable but does not follow a Poisson distribution. Therefore, we had to treat it as ordinal for the analyses (however, for descriptive statistics, we treated the variables as interval for illustrative purposes, as there are no analogs to means and standard deviations for ordinal data). A limitation of ordinal data methods is that each variable being compared (i.e. each combination of method and time) must have the same range of values. To accomplish this, we had to truncate the distribution of observed NW to integer values from zero to "four or more." The truncation affected 4.9% of observations.

Results

Descriptive statistics, including proportion of missing data relative to respondents who had any NW, WASO, latency, duration, or onset data at any time, are presented in Table 2 for NW and Table 3 for WASO. Means, standard deviations, and correlations in Table 3 were derived from the bootstrap, using the percentile bootstrap with 3,000 draws for significance tests.

NW analyses

Within-time analyses

Means. The means of actigraphy, diary, and BISQ number of night-wakings differed significantly in 2 degree-of-freedom (*df*) tests at all four time points [3-month: Wald $\chi^2(2, N = 199) = 93.28$, $p < .001$; 6-month: Wald $\chi^2(2) = 22.48$, $p < .001$; 12-month: Wald $\chi^2(2) = 24.04$, $p < .001$; 18-month: Wald $\chi^2(2) = 19.15$, $p < .001$). Pairwise comparisons within each time point (with FDR correction) were significant for 10 tests, with diary reports of NW trending higher than actigraphy and BISQ awakenings, except at 3 months when actigraphy showed the highest NW (Table 4—upper portion). Note that the means in Table 4 vary slightly from those in Tables 2 and 3 because of differences in the variables included in the model, in turn affecting the compensation for attrition. In no case does this variability affect substantive conclusions.

Correlations. Concomitant correlations between the three methods (actigraphy–diary, diary–BISQ, actigraphy–BISQ) were significant at all four time points, except for the correlation between actigraphy and the BISQ at 18 months, which was not significant (Table 2).

Concordance. We conducted concordance analyses for NW (treated as an ordinal variable as explained in the analysis plan)

Table 2. Descriptive statistics and correlations for NW across measure (actigraphy, sleep–diary, and BISQ) and time (3, 6, 12, and 18 months)

NW	Actigraph NW 3 m	Diary NW 3 m	BISQ NW 3 m	Actigraph NW 6 m	Diary NW 6 m	BISQ NW 6 m	Actigraph NW 12 m	Diary NW 12 m	BISQ NW 12 m	Actigraph NW 18 m	Diary NW 18 m	BISQ NW 18 m
ANW 3 m												
DNW 3 m	**0.58**											
BNW 3 m	**0.53**	**0.70**										
ANW 6 m	**0.42**	**0.40**	**0.42**									
DNW 6 m	**0.34**	**0.55**	**0.47**	**0.56**								
BNW 6 m	**0.37**	**0.55**	**0.44**	**0.48**	**0.76**							
ANW 12 m	0.10	0.10	−0.07	0.15	0.09	0.02						
DNW 12 m	0.13	0.36	**0.25**	0.09	**0.38**	**0.35**	**0.40**					
BNW 12 m	0.13	0.31	**0.32**	0.08	**0.30**	**0.41**	**0.27**	**0.68**				
ANW 18 m	0.12	0.19	−0.05	0.17	0.12	0.15	0.21	0.05	−0.02			
DNW 18 m	0.09	**0.22**	0.11	0.01	**0.21**	**0.23**	0.08	**0.38**	**0.30**	**0.27**		
BNW 18 m	0.11	**0.25**	0.16	0.07	**0.28**	**0.34**	0.10	**0.46**	**0.50**	0.06	**0.59**	
Mean	2.63	2.11	1.86	2.22	2.51	2.23	1.57	2.03	1.80	1.09	1.49	1.48
SD	1.07	1.18	1.08	1.23	1.24	1.22	1.02	1.20	0.38	0.89	0.99	1.11
Prop miss	0.15	0.07	0.09	0.23	0.14	0.18	0.37	0.26	0.28	0.50	0.42	0.41

Note: Full-sample $N = 226$ (prop miss uses $N = 226$ as the denominator). Variables treated as interval for purposes of table. Bold type indicates correlation is statistically significant, $p < .05$.

NW = number of nocturnal night-waking; m = months; ANW = actigraphic NW, DNW = diary NW; BNW = BISQ (Brief Infant Sleep Questionnaire) NW; prop miss = proportion missing.

Table 3. Descriptive statistics and correlations for nocturnal waking minutes (WASO) across measure (actigraphy, sleep diary, and BISQ) and time (3, 6, 12, and 18 months)

WASO	Actigraph WASO 3 m	Diary WASO 3 m	BISQ WASO 3 m	Actigraph WASO 6 m	Diary WASO 6 m	BISQ WASO 6 m	Actigraph WASO 12 m	Diary WASO 12 m	BISQ WASO 12 m	Actigraph WASO 18 m	Diary WASO 18 m	BISQ WASO 18 m
AWASO 3 m												
DWASO 3 m	**0.72**											
BWASO 3 m	**0.22**	**0.33**										
AWASO 6 m	**0.44**	**0.40**	0.00									
DWASO 6 m	**0.33**	**0.42**	**0.16**	**0.69**								
BWASO 6 m	0.10	**0.20**	**0.25**	0.15	**0.34**							
AWASO 12 m	0.12	0.12	0.09	**0.19**	0.01	0.09						
DWASO 12 m	0.05	0.21	**0.28**	0.10	0.17	**0.28**	**0.63**					
BWASO 12 m	**–0.19**	0.01	0.16	–0.09	0.00	**0.33**	**0.31**	**0.54**				
AWASO 18 m	0.06	0.09	0.04	**0.21**	0.02	0.04	**0.26**	0.25	0.00			
DWASO 18 m	0.04	0.18	0.13	0.17	0.16	0.07	**0.21**	**0.26**	0.10	**0.61**		
BWASO 18 m	0.03	0.18	0.22	0.03	0.05	**0.24**	0.10	**0.26**	**0.24**	**0.39**	**0.69**	
Mean	48.30	48.16	41.34	40.98	37.19	32.59	27.92	23.96	19.34	17.67	13.48	14.83
SD	24.59	29.36	51.09	24.68	28.40	34.96	19.61	22.10	34.09	13.41	13.72	18.42
Prop miss	0.15	0.16	0.12	0.22	0.16	0.20	0.37	0.29	0.31	0.49	0.40	0.43

Full-sample $N = 226$ (prop miss uses $N = 226$ as the denominator). Bold type indicates correlation is statistically significant, $p < .05$.
WASO = minutes awake after sleep; m = months; AWASO = Actigraphic WASO; DWASO = Diary WASO; BWASO = BISQ (Brief Infant Sleep Questionnaire) WASO; prop miss = proportion missing.

Table 4. Within-time mean comparisons between measures (actigraphy, sleep diary, BISQ)

NW	Actigraphy	Diary	BISQ
3 months	2.62[a]	2.11[b]	1.86[c]
6 months	2.20[a]	2.49[b]	2.22[a]
12 months	1.56[a]	2.02[b]	1.80[c]
18 months	1.03[a]	1.48[b]	1.44[b]
WASO	**Actigraphy**	**Diary**	**BISQ**
3 months	48.6[a]	48.7[a]	52.2[a]
6 months	41.1[a]	37.2[b]	32.7[b]
12 months	28.2[a]	24.2[b]	20.0[b]
18 months	17.2[a]	13.5[b]	15.1[ab]

N (night-wakings) = 201. N (WASO) = 201. Values within a row with the same superscript do not differ significantly, $p < .05$.
NW = number of nocturnal night-wakings; WASO = minutes awake after sleep onset; BISQ = Brief Infant Sleep Questionnaire.

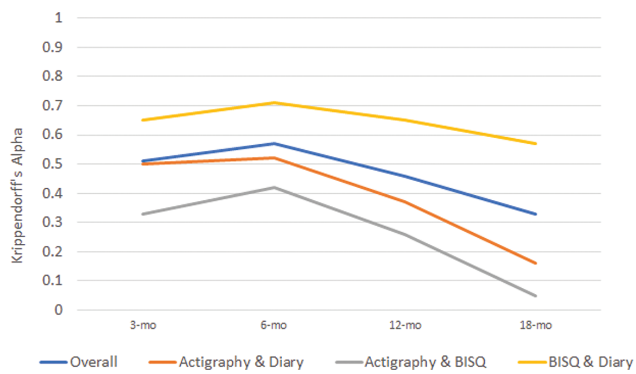

Figure 1. NW inter-method reliability.

using Krippendorff's α, a measure of consistency by different "raters" or, in this case, methods [46]. α has the advantage of consistent statistical underpinnings for different measurement types. We calculated alpha at each time across all three measurement types and between each pair. Krippendorff [46] recommends α > .80 as evidence for inter-method reliability and α > .667 as tentative evidence. Across all comparisons, the only estimates to reach even the lower threshold were between diary and BISQ ratings at 3, 6, and 12 months, with α = .65, .71, and .65, respectively. Figure 1 shows the pattern of all estimates. All comparisons show an apparent decline in concordance from 12 to 18 months, with the consistency between actigraphy and parent report measures at essentially chance levels at 18 months. Aggregated across all observed data, the diary measure of NW was greater than the actigraphy measure for 55% of nontied pairs and greater than the BISQ for 74%. Actigraphy NW was greater than BISQ for 62% of nontied pairs.

Across-time analyses

Correlations We assessed changes in the between-measure (actigraphy, diary, and BISQ) correlations across time using bootstrapping to generate confidence intervals (CIs) for the pairwise differences in correlations. There were six comparisons among the four time points for each pair of measures. In the absence of *p*-values (with asymmetric CIs), we used a simple adjustment of applying a two-tailed alpha of .01 (99% CI) for these tests. The correlation between actigraphy and the diary method at 18 months ($r = .27$) was significantly lower than that at 3 months ($r = .58$) and 6 months ($r = .56$). The same pattern was found for correlations between actigraphy and the BISQ (18 months: $r = .06$; 3 months: $r = .53$; 6 months: $r = .48$).

Trajectories We estimated the ordinal-data trajectory models for NW using a progression of models, relying on information criteria for model selection. For each of four models, Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and a sample-size-adjusted BIC (ssaBIC) are shown in Supplementary Table S-1. Lower values reflect better fit relative to parsimony. AIC and ssaBIC favored the quadratic model and BIC favored the latent basis model. We went forward with the quadratic model. The model-implied trajectories are shown in Figure 2. The trajectories are on an arbitrary scale as ordinal values do not have meaningful means, but the figure illustrates the relative shapes, and are interpretable. Parameter estimates are shown in Table 5.

We tested whether the linear and quadratic components of the latent trajectory, taken together, differed among the three measures of NW in a series of 2 *df* tests. We found that that the shape of the actigraphy trajectory differed from both the BISQ [$\chi^2(2) = 62.45$, $p < .001$] and the diary trajectories [$\chi^2(2) = 62.03$, $p < .001$], which did not significantly differ from each other [$\chi^2(2) = 3.95$, $p = .139$]. As shown in Figure 2, NW measured by actigraphy showed an approximately linear decline, while the BISQ and diary measures were approximately level from 3 to 12 months before declining from 12 to 18 months. All three assessment measures showed overall declines in NW from 3 to 18 months, *ps* < .001.

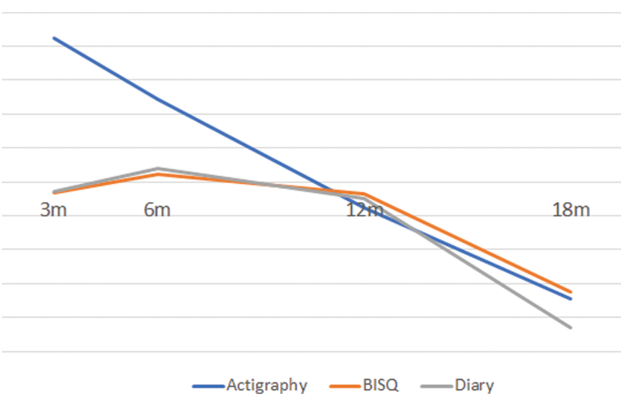

Figure 2. NW model-implied trajectories.

WASO analyses

Within-time analyses

Means. The means of actigraphy, diary, and BISQ WASO differed significantly in 2 *df* tests at 6, 12, and 18 months [6-month: Wald $\chi^2(2, N = 201) = 8.74$, $p = .013$; 12-month: $\chi^2(2) = 10.05$, $p = .007$; 18-month: $\chi^2(2) = 13.42$, $p = .001$)], but not at 3 months [$\chi^2(2) = 0.88$, $p = .644$]. Comparisons within each time point with FDR correction were significant for five tests, with a general pattern of actigraphy showing the highest WASO (Table 4—lower portion).

Correlations. Within-time correlations between the three methods (actigraphy–diary, diary–BISQ, actigraphy–BISQ) were significant at all four time points, except for the correlation between actigraphy and the BISQ at 6 months (Table 3). Overall, the correlations between actigraphy and the diary method were stronger than between actigraphy and the BISQ (for instance, at 3 months; actigraphy–diary: $r = .72$; actigraphy–BISQ; $r = .22$).

Concordance. To assess absolute agreement between methods in WASO, we generated the three pairwise Bland–Altman plots at each assessment point. The plots for 3-, 6-, 12-, and 18-month concordance between actigraphy and diary and between actigraphy and BISQ are shown in Figures 3, 4, 5, and 6, respectively. The left pane in each figure is the plot for actigraphy versus diary, and the right pane is actigraphy versus BISQ. Bland–Altman plots for concordance between diary and BISQ measures of WASO at each assessment point are in Supplementary Figure S-1. The horizontal axis in each plot is the "gold-standard" measure of WASO based on the mean of the two methods plotted. The vertical axis is the difference between measures, where positive values represent higher estimates from actigraphy than diary or BISQ and diary than BISQ. Deviation of the center line from zero indicates a nonzero mean difference between the two methods. The upper and lower bounds indicate the ±2 standard deviation (*SD*) bounds of the differences

Table 5. Descriptive statistics for night-waking latent trajectory parameters.

Parameter	Actigraphy intercept	Actigraphy linear slope	Actigraphy quadratic slope	Diary intercept	Diary linear slope	Diary quadratic slope	BISQ intercept	BISQ linear slope	BISQ quadratic slope
Actigraphy intercept									
Actigraphy linear S	−0.10								
Actigraphy quadratic S	**−0.74**	−0.12							
Diary intercept	**0.69**	−0.28	**−0.51**						
Diary linear S	−0.12	**0.84**	−0.20	−0.11					
Diary quadratic S	**−0.61**	−0.04	**0.92**	**−0.67**	−0.12				
BISQ intercept	**0.53**	**−0.36**	**−0.41**	**0.97**	−0.11	**−0.62**			
BISQ linear S	−0.03	**0.68**	−0.27	0.14	**0.96**	−0.25	0.16		
BISQ quadratic S	**−0.57**	−0.03	**0.83**	**−0.75**	−0.11	**0.98**	**−0.72**	−0.27	
Mean	N/A	**−2.56**	0.35	N/A	**−1.34**	**−2.03**	N/A	**−0.97**	**−1.58**
SD	1.77	1.43	3.08	2.74	2.12	3.60	2.74	2.06	3.09

N = 199. Means of trajectory intercept parameters have no scale for ordinal models. Bold type indicates parameter estimate is statistically significant, $p < .05$.
S = slope.

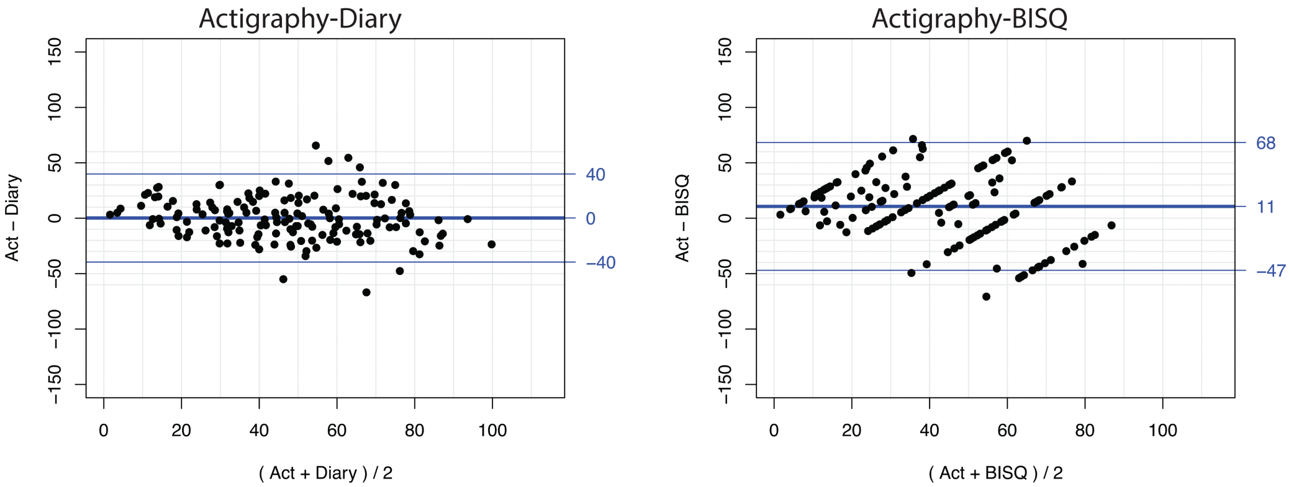

Figure 3. Plots of between methods difference against means for WASO at **3 months**. *Note*: The numbers on the left side of the y-axis are the difference between the two methods for the points plotted against the mean of the two methods (x-axis). The numbers on the right of the y-axis are mean (center) and ±2 *SD* of the difference. Plots for concordance between diary and BISQ measures of WASO at 3 months are in Supplementary Figure S-1.

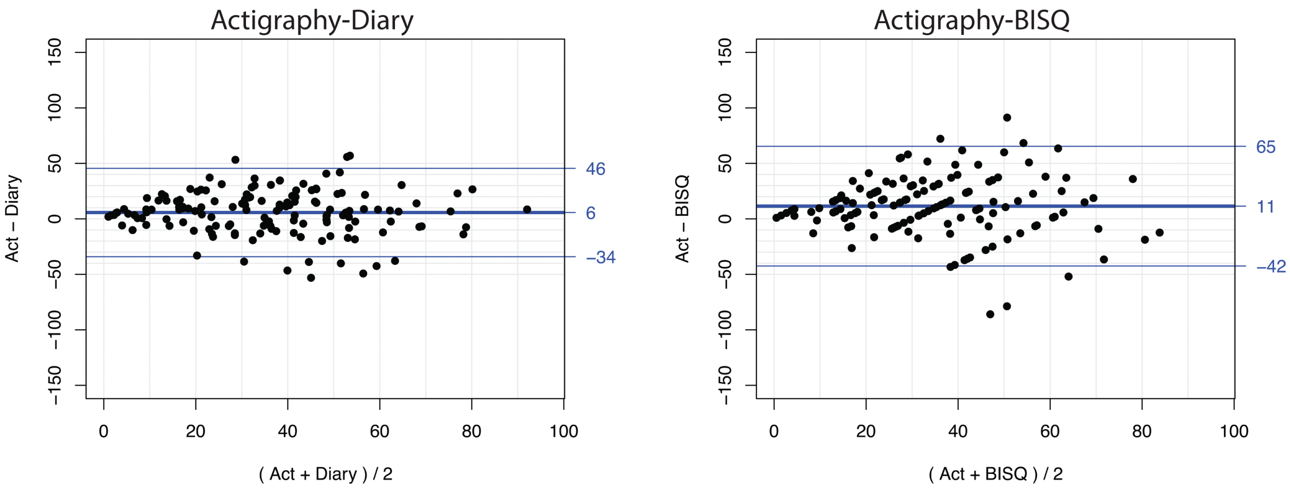

Figure 4. Plots of between methods difference against means for WASO at **6 months**. *Note*: The numbers on the left side of the y-axis are the difference between the two methods for the points plotted against the mean of the two methods (x-axis). The numbers on the right of the y-axis are mean (center) and ±2 *SD* of the difference. Plots for concordance between diary and BISQ measures of WASO at 6 months are in Supplementary Figure S-1.

("limits of agreement"), capturing approximately 95% of the observations.

At 3 months, actigraphy shows good aggregate concordance with diary, with a mean difference of zero. The bulk of the observations by the two methods are within 40 min of each other. Actigraphy averages 11 min greater than BISQ reports, with a wider confidence band (±57 min). Diary averages 7 min greater than BISQ, also with a wider confidence band of ±56.5 min. At 6 months, actigraphy averages 6 min greater than diary, with confidence bands of ±40 min. Actigraphy averages 11 min greater than BISQ reports, ±53.5 minutes. Diary averages 7 min greater than BISQ, ±45 min. At 12 months, actigraphy averages 6 min greater than diary, with confidence bands of ±29 min. Actigraphy averages 11 min greater than BISQ reports, ±38 min. Diary averages 5 min greater than BISQ, ±31.5 min. At 18 months, actigraphy averages 4 min greater than diary, with confidence bands of ±21.5 min. Actigraphy averages 6 min greater than BISQ reports, ±25 min. Diary averages 1 min greater than BISQ, ±20.5 min. Consistent with Werner et al.'s [12] limits of agreement and our own clinical experience, we a priori defined a satisfactory agreement if differences between the methods were smaller than 30 min. This prerequisite is attained for all three comparisons (actigraphy–diary, actigraphy–BISQ, diary–BISQ) only at 18 months. Over time, there is a trend toward narrowing of the confidence limits for each pairwise plot, which is expected in Bland–Altman plots of ratio variables as mean values and corresponding variances decline.

There is a general tendency for values to be greater for actigraphy than for diary, and for diary than for BISQ. Aggregated across all available pairwise comparison over assessment points, the actigraphic measure of WASO was greater than the diary measure for 62% of pairs, and greater than the BISQ for 71%. Diary WASO was greater than BISQ for 65% of pairs.

Across-time analyses

Correlations. We repeated the analysis of changes in the between-measure correlations across time. The only significant difference found (α = .01), was a greater correlation between the diary and the BISQ at 18 months (r = .69) than at 3 months (r = .33).

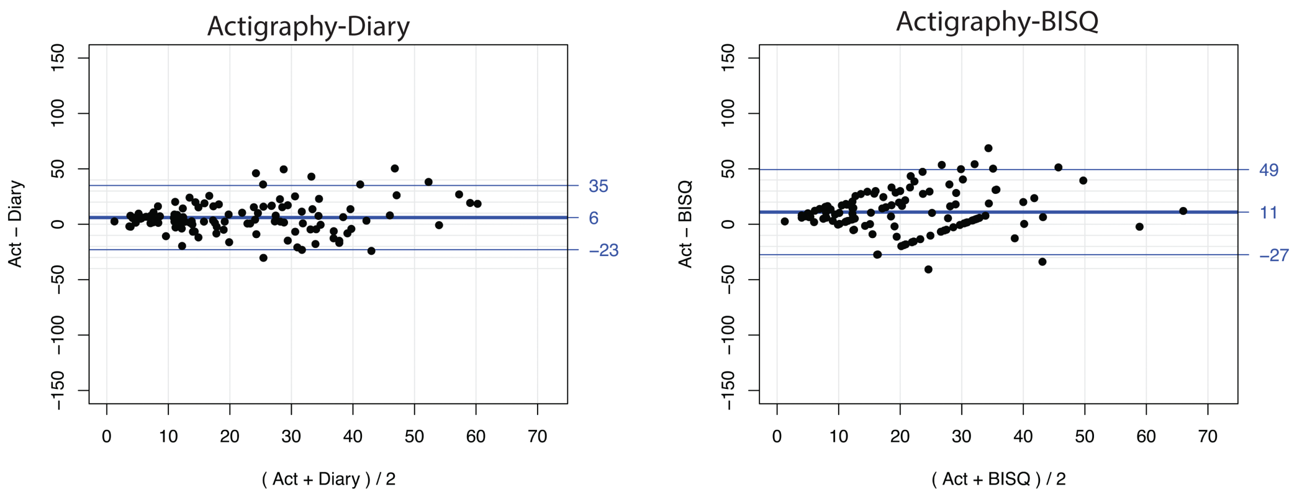

Figure 5. Plots of between methods difference against means for WASO at **12 months**. *Note*: The numbers on the left side of the y-axis are the difference between the two methods for the points plotted against the mean of the two methods (x-axis). The numbers on the right of the y-axis are mean (center) and ±2 *SD* of the difference. Plots for concordance between diary and BISQ measures of WASO at 12 months are in Supplementary Figure S-1.

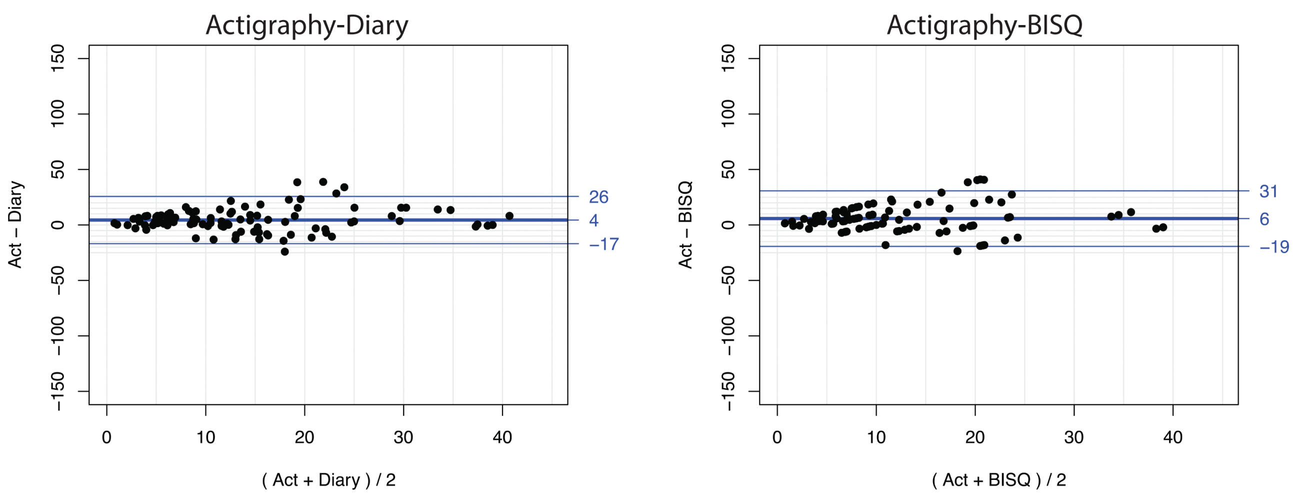

Figure 6. Plots of between methods difference against means for WASO at **18 months**. *Note*: The numbers on the left side of the y-axis are the difference between the two methods for the points plotted against the mean of the two methods (x-axis). The numbers on the right of the y-axis are mean (center) and ±2 *SD* of the difference. Plots for concordance between diary and BISQ measures of WASO at 18 months are in Supplementary Figure S-1.

Trajectories. In a model with all three WASO measures (actigraphy, diary, BISQ), information criteria favored the latent basis model (see Supplementary Table S-2 showing relative fit for trajectory models for WASO). However, this model (as well as the quadratic change model) resulted in several inadmissible estimates. Given that linear change was equally acceptable by BIC and fit well [$\chi^2(39) = 56.58$, $p = .034$, est. RMSEA = .047, 90% CI [0.014% to 0.073%], CFI = .97, TLI = .95, SRMR = .072] we proceeded with it. To improve stability of estimation, we constrained the small variance estimate of the BISQ linear slope to zero. This did not significantly impair fit, $\chi^2(6) = 10.69$, $p = .098$.

The model-implied trajectories appear in Figure 7. Actigraphy showed greater WASO across time (mean level) than the diary method, $M_{\text{diff}} = 2.72$, $SE = 1.02$, $z = 2.66$, FDR-adj. $p = .048$, but there were no other pairwise differences. Parameter estimates are shown in Table 6. All three measures showed overall declines in WASO from 3 to 18 months, *ps* < .001.

Additional measures

We created Bland–Altman plots and conducted latent trajectory analyses for three additional infant-sleep variables—sleep latency, sleep onset, and sleep duration—to complement the nighttime wakefulness analyses described above. Also for these variables we a priori defined a satisfactory agreement if differences between the methods were smaller than 30 min.

Sleep latency

Descriptive statistics for diary and BISQ sleep latency (based on the percentile bootstrap) are shown in Supplementary Table S-3. Bland–Altman plots for concordance between diary and BISQ measures of latency at each assessment point are in Supplementary Figure S-2. Average BISQ estimates were 5 min greater than diary reports at 3 months, with means by reporting method within one minute at 6, 12, and 18 months. Confidence bands ranged from ±26 min at 3 months to ±18–19 min at the

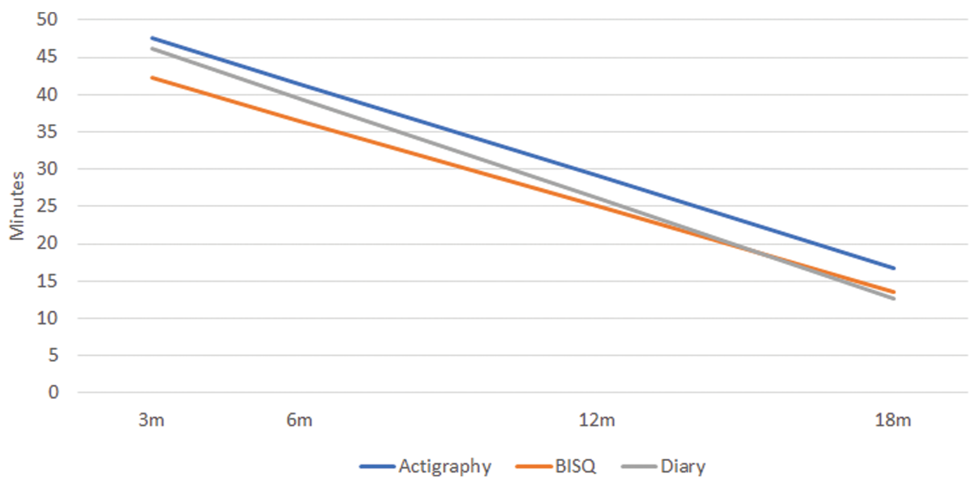

Figure 7. WASO model-implied trajectories.

three later assessments. The requirement of a difference smaller than 30 min was attained at all assessment points.

Among trajectory models, only the latent basis model fit acceptably without substantial impermissible estimates. Plotting the model-implied trajectories (Supplementary Figure S-3) showed both estimates to decline sharply from 3 to 6 months, and then gradually plateau. Trajectory parameter estimates are in Supplementary Table S-4.

Sleep duration

Descriptive statistics for actigraphy and diary sleep duration (based on the percentile bootstrap) are shown in Supplementary Table S-5. Bland–Altman plots for concordance between actigraphy and diary measures of sleep duration at each assessment point are in Supplementary Figure S-4. Mean diary reports were consistently greater than mean actigraphy values. At 3 months, diary reports exceeded actigraphy by 0.2 ± 1.3 h. At 6 months, the difference was 0.1 ± 0.65 h. At 12 months, the difference was 0.2 ± 0.85 h. At 18 months, the difference was 0.2 ± 0.8 h. The requirement of a difference smaller than 30 min was not achieved at any assessment point.

When considering LTMs, fit statistics supported either a latent trait (no change) or a linear change model (Supplementary Table S-5). The linear change model resulted in multiple impermissible estimates of correlations; we therefore selected the latent trait model which implied a constant value over time for both methods. We did not pursue more complex trajectories. The latent mean was significantly greater for the diary method, M = 10.41, SD = 0.54, than for the actigraphy method, M = 10.27, SD = 0.54; M_{diff} = 0.14, bootstrapped 95% CI [0.09% to 0.18%]. The correlation between the estimates was r = .84.

Sleep onset

Descriptive statistics for actigraphy, diary, and BISQ reports of sleep onset as measured in 24-h time are shown in Supplementary Table S-6. Onset times past midnight are reported as values greater than 24, for ease of calculation and comparison. Bland–Altman plots for concordance between actigraphy and diary (left panel), actigraphy and BISQ (center panel), and diary and BISQ (right panel) measures of sleep onset are shown for 3-, 6-, 12-, and 18-month measurements in Supplementary Figures S-5, S-6, S-7, and S-8, respectively. Mean actigraphy and diary reports were generally consistent with mean differences no more than 0.1 h. Confidence bands were ±1.2 h at 3 months and ranged from ±0.6 to ±0.8 h subsequently. Across assessment points, actigraphy mean values ranged from 0.3 to 0.5 h later than BISQ reports. Confidence bands of the difference declined consistently from ±1.95 h at 3 months to ±1.35 h at 6 months, ±1.1 h at 12 months, and ±0.95 h at 18 months. Mean diary reports were 0.3–0.4 h later than BISQ reports at every time. At 3 months, the confidence band of the difference was ±1.85 h, declining to ±0.95 to 1.15 h at subsequent occasions. The condition of a difference smaller than 30 min was not achieved at any assessment point.

None of the LTMs provided satisfactory fit with permissible solutions, indicating that there was no shape of the change over time that was consistent across families.

Discussion

This study examined for the first time the longitudinal correspondence between actigraphy, sleep diaries, and the BISQ in the assessment of infant nocturnal wakefulness (i.e. number and length of night-wakings) from 3 to 18 months postpartum. Another novel aspect of this study, which constitutes an important contribution to the literature on infant sleep consolidation, was the inclusion of trajectory models across development. These analyses allowed testing whether the trajectories differ in their growth components.

Infant nocturnal wakefulness

The trajectory analyses demonstrated overall declines for both WASO (i.e. minutes awake) and NW (i.e. number of night-wakings) from 3 to 18 months, according to actigraphy, sleep diaries, and the BISQ, suggesting that all three sleep assessment methods demonstrated a clear pattern of infant sleep consolidation over time [27, 29]. Nevertheless, some differences between the methods were apparent. For the NW measure, actigraphy showed an approximately linear decrease, while both reported methods demonstrated stability from 3 to 12 months and then decreased between 12 and 18 months. For WASO, though all three methods showed an overall linear decline, actigraphy had

Table 6. Descriptive statistics for WASO latent trajectory parameters

Parameter	Actigraphy intercept	Actigraphy linear slope	Diary intercept	Diary linear slope	BISQ intercept	BISQ linear slope
Actigraphy intercept						
Actigraphy linear S	**−0.48**					
Diary intercept	**0.73**	−0.41				
Diary linear S	−0.43	**1.14**	**−0.68**			
BISQ intercept	0.07	−0.02	**0.60**	**−0.36**		
BISQ linear S	−0.16	1.39	−2.52	2.42	**−3.59**	
Mean	**32.2**	**−20.6**	**29.5**	**−22.4**	**27.8**	**−19.1**
SD	10.4	12.3	11.6	12.1	15.3	2.3

N = 201. Bold type indicates parameter estimate is statistically significant, $p < .05$.
BISQ = Brief Infant Sleep Questionnaire, S = slope.

greater mean levels than parental reports over time. Similar differences in mean levels were also found when using within-time analysis. These findings are in line with previous studies demonstrating parental estimation of nighttime wakefulness to be shorter than actigraphy [10, 20, 27]. In the present study, actigraphic WASO was the highest at 6, 12, and 18 months, but at 3 months, no significant difference in mean levels was found between the three methods. Thus, a small difference in mean levels between the methods was evident mainly during the beginning of the infant's life, when the average time spent in nocturnal wakefulness was relatively long (~50 min). Prolonged awakenings at 3 months are likely to involve feeding or parental soothing and thus are naturally expected to be detected by all assessment methods.

A different pattern emerged with regards to the NW measure. While at 3 months, actigraphy showed the highest average number of NW, at 6, 12, and 18 months, NW were highest according to sleep diaries, followed by the BISQ. These findings (except at 3 months) are consistent with previous reports [29, 47], and probably result from the fact that our actigraphic NW measure referred to awakenings lasting 5 min or longer, whereas the parental reported NW measures consisted of all awakenings, including ones that were shorter than 5 min [10, 29].

Similarities and differences between the methods were apparent also when examining the bivariate correlations at each of the time points. Considering both WASO and NW, almost all of the concomitant correlations between the three methods were statistically significant at all assessment points. Moderate to high correlations were found between actigraphy and the diary (ranging between .27 and .58 for NW and .61 and .72 for WASO), and between the sleep diary and the BISQ. Thus, the diaries measures of nocturnal wakefulness, which share the parent-reported aspect with the BISQ and the day-to-day assessment with actigraphy, were associated with both the BISQ and the actigraph measures, whereas the lowest correlations were between actigraphy and the BISQ. This finding is not surprising, because the BISQ addresses global parental perceptions of infant sleep, whereas actigraphy represents an objective daily measure, which is based on motility [10, 29].

Although the correlation analysis demonstrates that the three assessment methods of nocturnal wakefulness are positively associated (overall, the more wakefulness detected by actigraphy, the more wakefulness parents report), this analysis does not fully assess the level of agreement between the methods [12, 33]. Using more rigorous measures of agreement revealed that the agreement between the methods in the assessment of nocturnal wakefulness, and especially between actigraphy and BISQ, was poor. We examined the agreement for NW, using Krippendorff's α and found evidence for tentative inter-method reliability only between the diary and the BISQ at 3, 6, and 12 months. Thus, even the two parent reported measures did not reach a high level of agreement (defined as α greater than .80). Agreement between actigraphy and the two parent reported measures of NW was low at all assessment points and reached a chance level at 18 months.

Consistent with the assessment of changes in the between-measure correlations for NW across time, the comparisons using Krippendorff's α showed an apparent decline in concordance between actigraphy and both parental methods with time, so that the agreement between methods was relatively higher at 3 and 6 months than at 18 months. This pattern of decline in the strength of the link between the objective and parental assessment methods may result from the growing ability of most infants after the age of 6 months to resume sleep following nocturnal awakening without calling for parental attention [10, 48]. Lower correspondence between parental reports and actigraphy in the number of awakenings is thus expected if, over time, the proportion of awakenings that go unnoticed by the parents increases.

Agreement rates for WASO were examined with Bland–Altman analysis. Following Werner et al.'s [12] limits of agreement and our own clinical experience, we defined a satisfactory agreement if differences between the methods were smaller than 30 min. Findings revealed that, although the mean differences between the methods were moderate (ranging from zero between actigraphy–diary at 3 months to 11 min between actigraphy–BISQ at 3 months), the limits of agreement at 3 and 6 months were above 30 min for all combinations (actigraphy–diary, actigraphy–BISQ, diary–BISQ), indicating poor agreement between the three methods in the assessment of nighttime wakefulness duration during early development. Similar to NW, the lowest agreement rates were between actigraphy and the BISQ with limits of agreement reaching 57 min at 3 months and 53 min at 6 months. Similar findings were reported before in toddlers and kindergarteners indicating poor agreement (with Bland–Altman) between actigraphy and the diary in the assessment of nocturnal wakefulness [11, 12]. In the current study, satisfactory agreement was found for WASO at 12 and 18 months (excluding actigraphy–BISQ at 12 months), emphasizing the importance of considering methods' concordance from a developmental perspective. Moreover, it seems that the satisfactory agreement rates at 12 and 18 months could mainly be explained by maturational processes leading to an increase in the proportion of infants who develop consolidated sleep (no or very short awakenings). As can be seen from the plots, for infants with relatively high WASO (above 20 min), there were still large differences between the methods even at 12 and 18 months. Because there is a general tendency for WASO to be greater for actigraphy than for both parental report methods, it seems that the disagreement results from parents being unaware of some of their infant's awakenings.

Additional measures

Although the focus of this study was on infant nocturnal wakefulness (i.e., number and length of NWs), we additionally examined the agreement between methods in the assessment of infant sleep latency, sleep onset, and sleep duration.

The diary and BISQ *sleep latency* measures (we did not have a measure of actigraphic sleep latency) demonstrated a high level of concordance along development. Both sleep latency trajectories showed a decline from 3 to 6 months and then stabilized. Concomitant correlations between the diary and BISQ were significant at all assessment points (ranging between .40 and .63), and agreement rates examined with Bland–Altman were satisfactory, indicating that the diary and the BISQ measures of sleep latency can be interchangeably used during infancy.

For *sleep onset* and *sleep duration*, the three methods demonstrated high concomitant correlations (ranging from .61 to .94) at all four assessment points. Although there were no large differences in mean levels between the three methods, the rate of agreement was not sufficient as the differences (limits of agreement) were larger than 30 min. This is consistent with Bélanger et al. [11], who also found poor agreement between actigraphy and diaries in the assessment of toddlers' sleep duration, but different from Werner et al. [12], who reported good agreement for sleep onset in kindergarteners.

In sum, although high associations were found between the methods in the assessment of sleep latency, sleep onset, and sleep duration, we found satisfactory agreement rates only for sleep latency, as assessed by the two parental-reported measures.

Limitations

This study has several limitations that merit consideration. Firstly, the integer nature of BISQ data for the number of NWs required some compromises in the analysis plan, including the truncation of NW reports to the smallest observed range. Further, our sample was relatively homogeneous, as it was comprised mainly of families from the middle-upper socioeconomic class in Israel. Finally, our sample included only first-born, normally developing infants. Although these characteristics may restrict the generalizability of the obtained results, we believe that these features of the sample enhance the internal validity of the study.

Conclusions

The current study demonstrated similarities and differences between actigraphy, sleep diaries, and the BISQ in the assessment of infant nocturnal wakefulness (i.e. number and length of NWs). The developmental trajectories of all three methods demonstrated a similar pattern of increase in sleep consolidation over time, and, at all assessment points, parental reports and actigraphy were significantly correlated. There were, however, also differences that should be taken into account. Most importantly, agreement between actigraphy and parental reports (especially the BISQ) in the assessment of the number of awakenings was poor at all assessment points and decreased along development, reaching a chance level at 18 months postpartum. This pattern probably suggests that parents are not aware of some of the infants' awakenings, and become even less so along development when more infants develop the ability to self-soothe. Agreement between methods in the assessment of the length of nocturnal wakefulness (WASO) improved overall along development but remained poor for infants with prolonged awakenings. Limited parental knowledge about infant awakenings could explain differences between actigraphy and parental reports, but is less relevant as an explanation for the gaps between the two parental measures, which showed good agreement only for sleep latency. Sleep diaries probably provide more accurate information on infant sleep than the BISQ, which represents a more general parental perception. This general perception is important when the focus of inquiry is the parents' subjective experience of their child's sleep. Thus, the decision on which method to use largely depends on the scientific or clinical question of interest. However, to achieve a comprehensive and accurate picture of infants' sleep and particularly of infant nighttime wakefulness, it is necessary to use both objective and subjective assessment methods. This seems to be especially important after the age of 6 months and for infants with suspected NW problems. Future studies should examine the convergence between the different sleep assessment methods in samples of clinically sleep-disturbed infants.

Supplementary material

Supplementary material is available at *SLEEP* online.

Acknowledgments

The authors wish to thank all participating families and all the students who helped with data collection. We thank Dr Gal Meiri for his help with handling the ethical approval process and for assisting with participants' recruitment. We thank Dr Patrick Malone for the statistical consultation. We are grateful to Prof. Avi Sadeh, our dear colleague, who sadly passed away, for his help and advice throughout the years. This was not an industry supported study.

Funding

This study was supported by a grant from the Israel Science Foundation (1075/10).
Conflict of interest statement. The authors have no conflicts of interest to disclose.

References

1. Bruni O, *et al*. Longitudinal study of sleep behavior in normal infants during the first year of life. *J Clin Sleep Med*. 2014;**10**(10):1119–1127.

2. Hysing M, *et al*. Trajectories and predictors of nocturnal awakenings and sleep duration in infants. *J Dev Behav Pediatr*. 2014;**35**(5):309–316.
3. Mindell JA, *et al*.; American Academy of Sleep Medicine. Behavioral treatment of bedtime problems and night wakings in infants and young children. *Sleep*. 2006;**29**(10):1263–1276.
4. Tikotzky L. Parenting and sleep in early childhood. *Curr Opin Psychol*. 2017;**15**:118–124.
5. Gregory AM, *et al*. Annual research review: sleep problems in childhood psychiatric disorders—a review of the latest science. *J Child Psychol Psychiatry*. 2016;**57**(3):296–317.
6. Bernier A, *et al*. Relations between physiological and cognitive regulatory systems: infant sleep regulation and subsequent executive functioning. *Child Dev*. 2010;**81**(6):1739–1752.
7. Byars KC, *et al*. Prevalence, patterns, and persistence of sleep problems in the first 3 years of life. *Pediatrics*. 2012;**129**(2):e276–e284.
8. Lam P, *et al*. Outcomes of infant sleep problems: a longitudinal study of sleep, behavior, and maternal well-being. *Pediatrics*. 2003;**111**(3):e203–e207.
9. Sadeh A. III. Sleep assessment methods. *Monogr Soc Res Child Dev*. 2015;**80**(1):33–48.
10. Sadeh A. A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample. *Pediatrics*. 2004;**113**(6):e570–e577.
11. Bélanger MÈ, *et al*. Investigating the convergence between actigraphy, maternal sleep diaries, and the child behavior checklist as measures of sleep in toddlers. *Front Psychiatry*. 2014;**5**:158.
12. Werner H, *et al*. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. *Arch Pediatr Adolesc Med*. 2008;**162**(4):350–358.
13. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev*. 2011;**15**(4):259–267.
14. Sadeh A, *et al*. The role of actigraphy in the evaluation of sleep disorders. *Sleep*. 1995;**18**(4):288–302.
15. 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.
16. 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.
17. 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.
18. Bélanger MÈ, *et al*. Validating actigraphy as a measure of sleep for preschool children. *J Clin Sleep Med*. 2013;**9**(7):701–706.
19. Meltzer LJ, *et al*. Use of actigraphy for assessment in pediatric sleep research. *Sleep Med Rev*. 2012;**16**(5):463–475.
20. Hall WA, *et al*. A comparison of actigraphy and sleep diaries for infants' sleep behavior. *Front Psychiatry*. 2015;**6**:19.
21. Sadeh A. Sleep assessment methods. In: El-Sheikh M, ed. *Sleep and Development: Familial and Socio-Cultural Consideration*. New York: Oxford University Press; 2011: 355–371.
22. Sadeh A. Evaluating night wakings in sleep-disturbed infants: a methodological study of parental reports and actigraphy. *Sleep*. 1996;**19**(10):757–762.
23. Sadeh A. Assessment of intervention for infant night waking: parental reports and activity-based home monitoring. *J Consult Clin Psychol*. 1994;**62**(1):63–68.
24. Anders TF, *et al*. Sleeping through the night: a developmental perspective. *Pediatrics*. 1992;**90**(4):554–560.
25. Sadeh A, *et al*. Parenting and infant sleep. *Sleep Med Rev*. 2010;**14**(2):89–96.
26. Tikotzky L, *et al*. Sleep patterns and sleep disruptions in kindergarten children. *J Clin Child Psychol*. 2001;**30**(4):581–591.
27. So K, *et al*. The use of actigraphy for assessment of the development of sleep/wake patterns in infants during the first 12 months of life. *J Sleep Res*. 2007;**16**(2):181–187.
28. Müller S, *et al*. 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.
29. Tikotzky L, *et al*. Maternal sleep-related cognitions and infant sleep: a longitudinal study from pregnancy through the 1st year. *Child Dev*. 2009;**80**(3):860–874.
30. Henderson JM, *et al*. The consolidation of infants' nocturnal sleep across the first year of life. *Sleep Med Rev*. 2011;**15**(4):211–220.
31. Teti DM, *et al*. Sleep arrangements, parent–infant sleep during the first year, and family functioning. *Dev Psychol*. 2016;**52**(8):1169–1181.
32. Mindell JA, *et al*. Sleep and social-emotional development in infants and toddlers. *J Clin Child Adolesc Psychol*. 2017;**46**(2):236–246.
33. Bland JM, *et al*. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;**1**(8476):307–310.
34. Mirmiran M, *et al*. Development of fetal and neonatal sleep and circadian rhythms. *Sleep Med Rev*. 2003;**7**(4):321–334.
35. de Weerd AW, *et al*. The development of sleep during the first months of life. *Sleep Med Rev*. 2003;**7**(2):179–191.
36. 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.
37. Sadeh A, *et al*. Activity-based sleep–wake identification: an empirical test of methodological issues. *Sleep*. 1994;**17**(3):201–207.
38. Muthén LK, *et al*. *Mplus User's Guide (v8.1)*. Los Angeles: Muthén & Muthén; 2018.
39. Bland JM, *et al*. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;**8**(2):135–160.
40. RC Team. *R: A Language and Environment for Statistical Computing (v.3.5.0)*. Vienna: Foundation for Statistical Computing; 2018.
41. Carstensen B, *et al*. Functions for Analysis of Agreement in Method Comparison Studies (v1.22.2). 2015. https://CRAN.R project.org/package=MethComp. Accessed August 10, 2018.
42. Martire RL. Reliability Coefficients (v.1.3.1). 2017. https://CRAN.R-project.org/package=rel. Accessed August 10, 2018.
43. Meredith W, *et al*. Latent curve analysis. *Psychometrika*. 1990;**55**(1):107–122.
44. Benjamini Y, *et al*. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met*. 1995;**57**(1):289–300.
45. Little RJA, *et al*. The analysis of social-science data with missing values. *Sociol Method Res*. 1989;**18**(2–3):292–326.
46. Krippendorff KT. *Content Analysis: An Introduction to Its Methodology* (2nd ed.) Thousand Oaks, CA: Sage; 2004.
47. Tikotzky L, *et al*. Sleep and physical growth in infants during the first 6 months. *J Sleep Res*. 2010;**19**(1 Pt 1):103–110.
48. Gradisar M, *et al*. Behavioral interventions for infant sleep problems: A randomized controlled trial. *Pediatrics*. 2016;**137**(6):e20151486.