Infant Behavior & Development 34 (2011) 541–551

Contents lists available at ScienceDirect

Infant Behavior and Development

Analyzing temporal patterns of infant sleep and negative affective behavior: A comparison between different statistical models

Mirja H. Hemmi[a,*], Silvia Schneider[b], Silvana Müller[a], Andrea H. Meyer[c], Frank H. Wilhelm[d]

[a] Swiss Etiological Study of Adjustment and Mental Health (sesam), Institute of Psychology, University of Basel, Switzerland
[b] Department of Clinical Child and Adolescent Psychology, Institute of Psychology, University of Bochum, Germany
[c] Department of Clinical Psychology, Division of Clinical Psychology and Psychotherapy, Section of Applied Statistics in Life Sciences, University of Basel, Switzerland
[d] Department of Clinical Psychology, Psychotherapy, and Health Psychology, University of Salzburg, Austria

ARTICLE INFO

Article history:
Received 26 October 2010
Received in revised form 30 May 2011
Accepted 23 June 2011

Keywords:
Infant sleeping
Negative affective behavior
Statistical modeling
Generalized additive mixed models

ABSTRACT

Objective: Variability in infant sleep and negative affective behavior (NAB) is a developmental phenomenon that has long been of interest to researchers and clinicians. However, analyses and delineation of such temporal patterns were often limited to basic statistical approaches, which may prevent adequate identification of meaningful variation within these patterns. Modern statistical procedures such as additive models may detect specific patterns of temporal variation in infant behavior more effectively.
Method: Hundred and twenty-one mothers were asked to record different behaviors of their 4–44 weeks old healthy infants by diaries for three days consecutively. Circadian patterns as well as individual trajectories and day-to-day variability of infant sleep and NAB were modeled with generalized linear models (GLMs) including a linear and quadratic polynomial for time, a GLM with a polynomial of the 8th order, a GLM with a harmonic function, a generalized linear mixed model (GLMM) with a polynomial of the 8th order, a generalized additive model, and a generalized additive mixed model (GAMM).
Results: The semi-parametric model GAMM was found to fit the data of infant sleep better than any other parametric model used. GLMM with a polynomial of the 8th order and GAMM modeled temporal patterns of infant NAB equally well, although the GLMM exhibited a slightly better model fit while GAMM was easier to interpret. Besides the well-known evening clustering in infant NAB we found a significant second peak in NAB around midday that was not affected by the constant decline in the amounts of NAB across the 3-day study period.
Conclusion: Using advanced statistical procedures (GAMM and GLMM) even small variations and phenomena in infant behavior can be reliably detected. Future studies investigating variability and temporal patterns in infant variables may benefit from these statistical approaches.

Abbreviations: NAB, negative affective behavior; GLM, generalized linear models; GLMM, generalized linear mixed models; GAM, generalized additive models; GAMM, generalized additive mixed models.

* Corresponding author at: Swiss Etiological Study of Adjustment and Mental Health (sesam), National Centre of Competence in Research, Birmannsgasse 8, CH-4009 Basel, Switzerland. Tel.: +41 61 267 02 22; fax: +41 61 267 02 74.
E-mail address: mirja.hemmi@unibas.ch (M.H. Hemmi).

0163-6383/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.infbeh.2011.06.010

1. Introduction

The first months of an infant's life are a period of adaptation and low behavioral stability characterized by high intra- and interindividual variability in the amounts of sleep as well as crying or fussing behaviors during wake times (Belsky, Fish, & Isabella, 1991; de Weerth & van Geert, 2002; Emde, Gaensbauer, & Harmon, 1976; Fish, Stifter, & Belsky, 1991). While sleep serves a basic homeostatic function important for regeneration, growth and neural development (Davis, Parker, & Montgomery, 2004), crying and fussing can be regarded as early behavioral indicators of a negative affective state of the infant (Lummaa, Vuorisalo, Barr, & Lehtonen, 1998). The variation of prevalence and diurnal pattern of these infant behaviors has long been of interest to researchers and clinicians (Bamford et al., 1990; Barr, 1990; Brazelton, 1962; St. James-Roberts & Halil, 1991). Sudden shifts and high variability in infant sleep and negative affective behavior (NAB) may easily concern many parents resulting in frequent complaint and family disruption (Forsyth, Leventhal, & McCarthy, 1985; St. James-Roberts & Halil, 1991). Thus, the distinction between variation in infant sleep and NAB as either a normal developmental phenomenon or as a clinical problem is of major importance (Barr, 1990; Brazelton, 1962).

To date, infant diaries are the simplest and most cost-effective technique to record infant behavior (Lam et al., 2010). Diaries as well as other highly sensitive time-series measurement methods (e.g., actigraphy) typically provide large numbers of serially connected data points for each infant. These data allow both the detection of between-subject differences and of within-subject daily fluctuations in circadian sleep and NAB patterns. Hence, care must be taken to find appropriate statistical models capable of handling the abundance of data for these behavioral outcomes.

Since the early 1970s generalized linear models (GLMs) have been used for modeling infant sleep and NAB (McCullagh & Nelder, 1983; Nelder & Wedderburn, 1972). GLMs are mathematical extensions of linear models with the ability to handle a larger class of distributions of outcome variables. They have the advantage of being more flexible while allowing for non-normally distributed errors and/or non-constant variance structure in the data (Guisan, Edwards, & Hastie, 2002; Hastie & Tibshirani, 1990). In the present study infant sleep and NAB followed a binomial error structure for which a GLM is suitable in principle. A major limitation of the GLM, however, is the assumption of independence of observations. Serial dependency of data points is usually found in diary-captured behavior data and a violation of this assumption leads to an inflation of Type I errors (Holditch-Davis, Edwards, & Helms, 1998; Zuur, Ieno, Walker, Saveliev, & Smith, 2009). For this reason, the generalized linear mixed model (GLMM), an extension of the GLM, is a more suitable approach. GLMMs allow for correlation between repeated observations and for nested data structures to model temporal circadian patterns in repeated measures designs (Holditch-Davis et al., 1998; Zuur et al., 2009). They may also include time-invariant (trait-specific) as well as time-dependent (state-specific) explanatory variables (Singer & Willett, 2003). This provides a solution to the important aim of identifying inter-individual differences in intra-individual change (Baltes & Nesselroade, 1979; Hoeksma & Koomen, 1992) i.e., the questions of whether temporal trends of groups of individuals vary from each other. A major disadvantage of GLMs and GLMMs is their limitation in predicting non-linear responses over time, which is, however, an inherent characteristics of infant sleep and NAB (Cudeck, 1996).

Semi-parametric regression models (i.e., extensions of GLMs), such as generalized additive models (GAMs), use a different approach (Faraway, 2006; Hastie & Tibshirani, 1990). These models allow for non-linear relationships between the response variable and multiple explanatory variables using smoothing functions (Zuur et al., 2009). GAMs are flexible and effective techniques for conducting nonlinear regression analyses, for example in longitudinal studies in which patterns of temporal trends are to be estimated (Dominici, McDermott, Zeger, & Samet, 2002; Guisan et al., 2002). The linear function of the explanatory variable, i.e., the variable denoting time, is thereby replaced by an additive function (Hastie & Tibshirani, 1986). The major advantage of the GAM is its ability to deal with highly non-linear and non-monotonic relationships between the response and the set of explanatory variables (Guisan et al., 2002). These models are data rather than model driven and it is therefore the data that determine the nature of the relationship between the response and the set of explanatory variables, thereby not assuming some form of parametric relationship. This makes them a highly flexible tool to estimate complex temporal patterns (Guisan et al., 2002; Yee & Mitchell, 1991). Generalized additive mixed models (GAMMs), an extension of GAMs, work on a similar level as GLMMs in GLMs. They additionally take the auto-correlation among repeated measures into account by including random effects and thus distinguish between-subject from within-subject variability (Wood, 2006).

To date, the majority of studies that investigated developmental and circadian trajectories of infant sleep and NAB are based on smoothed patterns with aggregated data (Barr, 1990; Brazelton, 1962; St. James-Roberts & Halil, 1991). Studies that investigated individual variability and trajectories per se are scarce and usually limited to the developmental course of either sleep or NAB (Bamford et al., 1990; de Weerth & van Geert, 2002; Holditch-Davis et al., 1998; Iglowstein, Jenni, Molinari, & Largo, 2003; St. James-Roberts & Plewis, 1996; van den Boom & Hoeksma, 1994; van Geert & van Dijk, 2002). Although these studies captured infant sleep or NAB by sensitive measurement methods (e.g., diaries or observations) and assessed variability predominantly by parametric (GLMM) and semi-parametric (GAM) analyses, there are various methodological concerns. Two studies (de Weerth & van Geert, 2002; van Geert & van Dijk, 2002) were based on very small sample sizes. Short observation periods in two further studies (de Weerth & van Geert, 2002; Holditch-Davis et al., 1998) did not allow detection of variability in 24 h patterns. Harmonic analyses, which were conducted in one study (Bamford et al., 1990) accounted for the circadian rhythm of infant sleep, however, were unable to integrate the repeated nature of diary data and therefore lacked the identification of individual trajectories. Only one study analyzed infant sleep by the semi-parametric procedure GAM. However, data on infant sleep were collected from parents only by structured interviews concerning various sleep-related habits from infancy to adolescence (Iglowstein et al., 2003). One of the rare studies to investigate daily fluctuations in infant

sleep and NAB (St. James-Roberts & Plewis, 1996) employed a GLMM approach, but used total amounts of infant sleep and NAB rather than individual data points for analysis. To our knowledge, despite the advantage of additive models no study has analyzed infant sleep or NAB captured by high temporal resolution repeated measurement methods with GAMMs so far nor is there any published article that modeled infant behavior using different statistical approaches simultaneously in the same study sample for direct comparison. This may in part be because GAMM is still on the frontier of statistical research and available documentation is rather technical (Zuur et al., 2009).

The purpose of this paper is to compare four different parametric models with the two semi-parametric procedures GAM and GAMM for modeling circadian probability patterns of sleep and NAB in healthy infants aged 4–44 weeks. Our aim was to demonstrate the usefulness of additive models as a statistical tool to describe and to test temporal courses of infant sleep and infant NAB. Amounts of infant sleep, waking, and NAB were collected with infant diaries by mothers during three consecutive days. Our principal interest was on circadian pattern as well as day-to-day variability in infant sleep and NAB. In order to make a fair comparison of the parametric with the two additive models we had to employ a relatively high order polynomial that was able to fit the observed values reasonably well. Hence, the following parametric and semi-parametric models were used to model infant sleep and NAB: GLMs including a linear and quadratic polynomial for time, a GLM with a polynomial of 8th order, a GLM with a harmonic function, a GLMM with a polynomial of 8th order, a GAM, and a GAMM. We expected to find infant behavior patterns that have previously been reported in the literature, including the NAB peak in the evening (Barr, 1990; St. James-Roberts & Halil, 1991) and the typical sleep pattern, including daytime sleep in the afternoon (Bamford et al., 1990; Iglowstein et al., 2003). However, we also assumed that due to the unique properties of the advanced models we would be able to detect infant behavior phenomena that have not been reported before. We hypothesized that both semi-parametric approaches, GAM and GAMM, would perform best in modeling the complex circadian pattern of infant sleep and NAB.

2. Methods

This study was part of the Swiss Etiological Study of Adjustment and Mental Health (sesam), an epidemiologic study investigating psychological, environmental, and biological influences on the development and mental health of children in Switzerland. The aim was to assess characteristics of infant crying, feeding, and sleeping behavior by infant diaries, actigraphy, and interview. In addition, maternal psychological trait characteristics were examined by questionnaires. This report exclusively focuses on the infant diary data.

2.1. Participants and procedures

Hundred and sixty-seven mothers with healthy infants aged 4–44 weeks and with sufficient knowledge of the German language responded to advertisements on the Internet or were approached through day nurseries, parent counseling centers, pediatricians, and midwifes in Switzerland and Germany. No attempt was made to select by social class. Mothers were asked to participate in a study about infant sleeping, crying, and feeding behavior by filling out behavior diaries for three consecutive days. Twenty-four mothers did not want to participate after initial contact, 18 were unreachable, and four were excluded for organizational reasons, leaving 121 mothers in the final sample. Diary data of 116 (96%) infants were successfully collected over the 3-day period. The missing data were due to early withdrawal from the study ($n = 2$) or not sending back the diaries ($n = 3$). Table 1 shows characteristics of the final sample. Average age of mothers was approximately 32 years and infants were on average approximately 20 weeks old. Almost all diaries (99.1%) were completed for the entire study period and contained very few missing data.

Diary and instructions were mailed together with a stamped addressed return envelope. Mothers filled in the diaries for three consecutive days (72 h period), starting all at the same weekday (Tuesday) in order to obtain a standardized picture of infants' common activities during working days. Detailed information on how to fill in the diaries was provided to mothers by written instructions and by a scheduled telephone session. They were asked to fill in the diaries relatively often on a regular basis during the day, but at convenient times in order to not interrupt their normal daily activities. All infant and caregiver symbols were explained and special descriptions for the distinction of crying, fussing, and unsoothable crying were provided. Fussing was defined as "Baby makes unhappy sounds without crying" and unsoothable crying was defined as "Baby is crying without any success to console it."

The study was approved by the medical ethics committee of Basel, Switzerland, and all participating mothers gave their written informed consent.

2.2. The Baby's Day Diary

To obtain a detailed picture of an infant's common activities during a 24 h period, the paper-and-pencil version of the Baby's Day Diary (Barr, Kramer, Boisjoly, McVey-White, & Pless, 1988) was used, which our research group translated into German. This widely adopted measurement method provides a simple and cost-effective technique to record the most common infant behaviors (i.e., sleeping, being awake and content, fussing, crying, unsoothable crying, and feeding) as well as two caregiver behavior categories directly relevant to the infant (i.e., body contact/carrying/holding and moving by baby carriage or car). The Baby's Day Diary is a letter-sized sheet containing four time-rulers arranged vertically, each indicating

Table 1
Characteristics of the sample.

	Mean (SD) or %
Number of infants assessed	116
Infant age (weeks)	20.4 (7.9)
Maternal age	32.0 (4.1)
Paternal age	34.2 (5.3)
Proportion of boys (%)	53.5
Proportion	
Firstborns (%)	67.1
Secondborns (%)	23.9
Thirdborns (%)	6.3
Fourthborns (%)	2.7
Number of years of education of the mother	16.7 (3.3)
Proportion of married or living in a relationship (%)	90.9
Proportion of monthly household income (%, CHF)	
<2500	11.0
2500–5000	22.3
5000–7500	36.6
>7500	30.2
Proportion of three completed diary days (%)	99.1
Proportion of missing data (%)	0.6
Number of minutes of infant behavior per 24-h	
Sleeping	787 (91)
Awake and content	403 (103)
Negative affective behavior (NAB)	101 (61)

one period of the day:night (midnight to 6 am), morning (6 am to noon), afternoon (noon to 6 pm), and evening (6 pm to midnight). Each time-ruler is divided into a baby and a caregiver section. The temporal resolution of the 24 h period is in 5-min intervals. Using the line styles and symbols assigned to each behavior, indicated on a legend on the sheet, caregivers recorded the onset and end time of each infant behavior.

The Baby's Day Diary has been widely used and validated against objective measurement methods (Barr et al., 1988; Barr, Kramer, Pless, Boisjoly, & Leduc, 1989; Hunziker & Barr, 1986; St. James-Roberts, Hurry, & Bowyer, 1993; St. James-Roberts & Plewis, 1996). Similar cry durations have been reported for maternal diary report and coded audio-recordings (St. James-Roberts et al., 1993). Over 24 h, significant medium to large correlations were found for sleep duration between an electronic version of the Baby's Day Diary and actigraphy ($r = 0.41–0.65$) and between the paper-and-pencil diary version we used in our study and actigraphy ($r = 0.47–0.70$) (Müller, Hemmi, Wilhelm, Barr, & Schneider, 2011). The high temporal resolution of 5 min has been shown to more accurately capture crying bouts during development than a lower resolution (Barr et al., 1988; Miller, Barr, & Eaton, 1993).

2.2.1. Coding of infant sleep/wake cycle and negative affective behavior (NAB)

Infant sleep/wake cycle was measured continuously over 72 h. The variable was dichotomous with 0 indicating the infant's waking state (which may be awake and content, crying, fussing, feeding) and 1 indicating the infant's sleeping state. For the analyses of NAB, infant crying and fussing were collapsed since initial analyses indicated that the frequency of infant crying was relatively low in our non-clinical sample (mean amount of total cry duration during 24 h:25.2 (±22.8) min). Thus, following previously used procedures (St. James-Roberts, Conroy, & Wilsher, 1998), NAB was composed of the three diary codes fussing, crying, and unsoothable crying and in the analyses it was exclusively contrasted to the infants' common positive affective state (i.e., awake and content). Times when mothers indicated that the infant slept or was fed were excluded in these analyses because no unambiguous information of the infants' affective state can be obtained during sleeping and feeding situations. To find a suitable time range for reliable statistical analysis of NAB, diary data were plotted against time of day for the entire sample and the number of data points for NAB or awake and content codes was examined for every hour. Since sleep was prevalent between 8 pm and 8 am (more than 70% of data points) the time range 8 am to 8 pm was selected for the analysis of NAB. Infant NAB was dichotomous with 1 indicating that the infant was awake and in a distressed state showing NAB, whereas 0 indicated that the infant was awake and in a content state.

2.3. Statistical analysis

We modeled the probability of sleep and NAB within and across the three days with a temporal resolution of five minutes resulting in a maximum of 288 time points per day or 864 time points across the three days for each participant resulting in 100,224 data points across all 116 infants during the study period. Of these, 99,648 (99.4%) valid data points were included in the analysis of infant sleep (55% asleep vs. 45% awake). For the analysis of infant NAB where the time range was restricted to 8 am to 8 pm and data points regarding infant sleep and feeding behavior were excluded during this period, a total of 29,954 of the maximum possible 35,029 data points (85.5%) were included (82% awake and content vs. 18% NAB). The first model was a GLM with a linear and a quadratic polynomial for the explanatory variable denoting time and with the probability

of sleep or NAB as response variables (Model 1). This model can be seen as a relatively simple model that accounts for a curvilinear shape of the daily temporal pattern as was expected, especially for the probability of sleep. In order to make a fair comparison of the parametric with the additive models we had to employ a GLM with a polynomial of 8th order (including all lower-order terms; Model 2). To obtain this model we successively increased the number of polynomials until the inclusion of additional polynomials did not significantly improve model fit. As an additional model for the probability of sleep a harmonic regression function with three superimposed sine and cosine frequencies was used (Model 3), which is suitable for outcomes with a circadian rhythm (Bamford et al., 1990). In a second step, a generalized linear mixed model (GLMM) with a polynomial of the 8th order and with a random intercept (Model 4) was included (Littell, Miliken, Stroup, Wofinger, & Schabenberger, 2006). GLMMs account for the serial dependence of the data points, caused by the repeated measures structure of the data, by including a random coefficient for individuals. In a final step, infant sleep and NAB were analyzed with the semi-parametric models GAM (Model 5) and GAMM (Model 6). The amount of smoothing was computed using penalized regression splines (Wood, 2006). Penalization was necessary to protect against overfitting of the data leading to unnecessary complexity of splines. Cubic regression splines were used in which cubic polynomial functions were fitted within different intervals of the predictor variable and smoothly connected across intervals. The amount of smoothing was not fixed to a preset value but determined using cross validation (Wood, 2006). Cyclic cubic regression splines were used for the circadian rhythm pattern of infant sleep probability in order to avoid a gap between the end and onset of different days. As a null model (Model 0) we included a model that assumes constant and equal probabilities of sleep or NAB over time for each subject.

The different models can be written as follows:

Model 0: $\log(p_{is}/1-p_{is}) = \beta_0$
Model 1: $\log(p_{is}/1-p_{is}) = \beta_0 + \beta_1 \text{hours}_{is} + \beta_2 \text{hours}_{is}^2$
Model 2: $\log(p_{is}/1-p_{is}) = \beta_0 + \beta_1 \text{hours}_{is} + \beta_2 \text{hours}_{is}^2 + \cdots + \beta_8 \text{hours}_{is}^8$
Model 3: $\log(p_{is}/1-p_{is}) = \beta_0 + \sin(2\pi/24)\text{hours}_{is} + \cos(2\pi/24)\text{hours}_{is} + \sin(2\pi/12)\text{hours}_{is} + \cos(2\pi/12)\text{hours}_{is}$
$+ \sin(2\pi/8)\text{hours}_{is} + \cos(2\pi/8)\text{hours}_{is}$
Model 4: $\log(p_{is}/1-p_{is}) = \beta_0 + \beta_1 \text{hours}_{is} + b_s$
Model 5: $\log(p_{is}/1-p_{is}) = \beta_0 + f(\text{hours}_{is})$
Model 6: $\log(p_{is}/1-p_{is}) = \beta_0 + f(\text{hours}_{is}) + b_s$

where p_{is} = probability of sleep/NAB for a subject s at time point i, hours = time of day in hours ranging from 0 to 24, β_0 = intercept; β_1–β_8 = regression coefficients of time; $f(\,)$ = smoothing function (here of type penalized cubic regression spline); b_s = individual random effect where $b_s \sim \mathrm{N}(0, \sigma_b^2)$; $y_{is} \sim B(1, p_{is})$ where y_{is} takes values 0 or 1.

For comparisons among different models, day-to-day variability was not considered. However, we ran a final analysis of the best fitting model, which included the terms day and the interaction term of time of day and day to test for variation in daily courses among the three days.

The model for the analysis of daily fluctuations can be written as follows:

$$\log \frac{p_{is}}{1-p_{is}} = \beta_0 + f_j(\text{hours}_{is}) + \text{day}_{js} + b_s$$

where p_{is} = probability of sleep/NAB for a subject s at time point i, hours = time of day in hours ranging from 0 to 24, β_0 = intercept; $f_j(\,)$ = smoothing function for a day j (here of type penalized cubic regression spline); b_s = individual random effect where $b_s \sim N(0, \sigma_b^2)$; $y_{is} \sim B(1, p_{is})$ where y_{is} takes values 0 or 1.

For all model computations the software R (R Development Core Team, 2009) was used. For the computations of GAMs and GAMMs the packages lme4 (Bates & Maechler, 2009) and gamm4 (Wood, 2009) were used. The function gamm4, which is found in the R package of the same name, is numerically more robust than the function gamm of the R package mgcv for dichotomous data like those presented here (Wood, 2009). The function gamm4 also allows model comparison on the basis of the log likelihood or information criteria (Bayesian Information Coefficient, BIC). Goodness of fit was based on the BIC with smaller values denoting better fit. This index takes model complexity into account by including a penalty term for the number of parameters and the sample size used.

3. Results

3.1. Comparison of the models for temporal patterns of infant sleep

Fig. 1a shows the 24 h probability pattern of infant sleep for the observed data. The pattern follows the characteristic shape of a circadian rhythm for older infants. Daytime activities start around 8 am and last until 8 pm with temporarily increased sleep probabilities around the early afternoon relating to post-lunch daytime napping.

Table 2 displays model fit indices for the different analyses of infant sleep. A comparison among these models revealed that the 2nd order polynomial model (1) which describes a simple curvilinear shape, fitted the data poorly as it neglected the midday peak (Fig. 1b). A GLM with a polynomial of the 8th order (Model 2) fitted the data better than Model 1 since it also allowed for the small peak in the afternoon. A GLM with a harmonic function (Model 3) was the first model that accounted

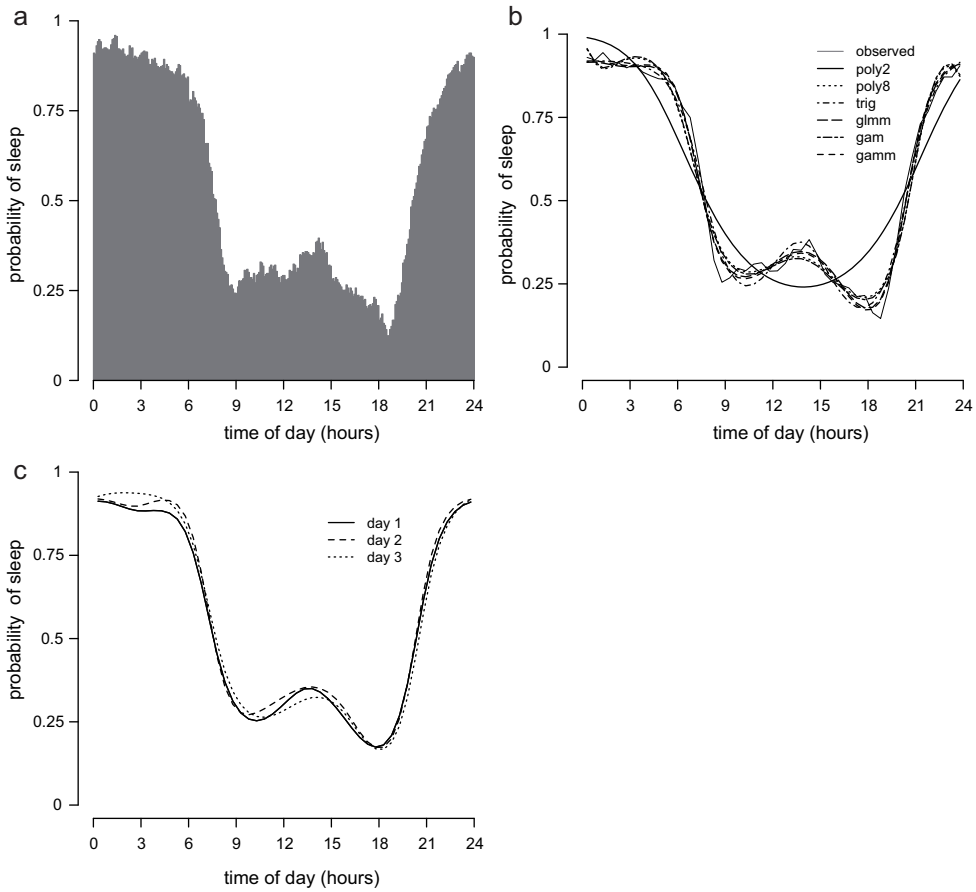

Fig. 1. (a) Spine plot showing the daily course of observed probabilities of infant sleep. (b) Daily course of predicted values of different models for probability of infant sleep. (c). Daily course of predicted values for probability of infant sleep for three consecutive days.

for both the afternoon peak in the probabilities of sleep and a smoothed transition between the end and onset of two days. However, the harmonic model had a model fit less adequate than the GAM (Model 4) and the two mixed models. A GLMM with a polynomial of the 8th order and with a random intercept (Model 5) exhibited AIC values close to the GAMM (Model 6), the best fitting model for predicting probabilities of infant sleep. GLMs and the GLMMs both suffered from an evident gap between the transitions of two days in the predicted values (Fig. 1b). Additive models generally predicted the shape of the temporal curve better, especially the GAMM. The temporal pattern modeled by the GAMM clearly reveals the circadian rhythm of the probability of sleep while underlining the small increase in the early afternoon and a smoothed transition between the end and onset of two days. Using GAMM, highest probabilities for sleep were obtained at midnight with sleep probabilities of 96%, whereas lowest probabilities were found around 10 am and 4 pm with probabilities between 25% and 17%, respectively. The early afternoon peak reached a probability value of 34%.

The GAMM model that included in addition day as predictor revealed a significant interaction effect between time and study day and a main effect of the study day variable (Table 3 and Fig. 1c). However, difference in percentage points between the three days were rather small with a maximum of approximately six percentage points between the first and the third day at 3 am (Fig. 1c).

Table 2
Model fits for the probability of sleep.

Model	logLik ($\times 10^4$)	BIC ($\times 10^4$)	Parameters	Deviance explained (%)
0 GLM null model	−6.8627	13.7266	1	
1 GLM polynomial 2	−5.3180	10.6395	3	22.5
2 GLM polynomial 8	−5.1110	10.2325	9	25.5
3 GLM trigonometric	−5.0902	10.1885	7	25.8
4 GAM	−5.0846	10.1650	2	25.9
5 GLMM	−5.0270	10.0657	10	26.7
6 GAMM	−4.9997	10.0035	3	27.1

Table 3
Model fits for probability of sleep: main and interaction effects for time of day and study day using a GAMM.

	logLik ($\times 10^4$)	BIC ($\times 10^4$)	Parameters	Deviance change	Change in parameters	*p*-Value
Sleep						
GAMM Intercept + time	−4.9997	10.0035	3			
GAMM Intercept + time + day	−4.9986	10.0029	5	22.8	2	<0.001
GAMM Intercept + time + day + time × day	−4.9967	10.0003	7	37.2	2	<0.001

GAMM = generalized additive model with a random intercept; parameter = number of estimated parameters; logLik = log-likelihood of model; deviance change = difference in −2 × logLik between model and previous model; change in parameters = difference in number of estimated parameters between model and previous model; BIC = Bayesian information criterion; time = 24 h period; day = study day.

3.2. Comparison of the models for temporal patterns of infant NAB

Fig. 2a shows the 12 h probability pattern of infant NAB for the observed data. The probability of infant NAB increased during the day, though not monotonically as seen by the two daily peaks, the first around midday and the second in the early evening, a time often associated with night sleep onset.

Table 4 shows all model fit indices for the different models of infant NAB. Due to the non-curvilinear shape of the probability of infant NAB, the 2nd order polynomial baseline model (Model 1) fitted the data poorly and on a similar level as the null model (Model 0). A GLM with a polynomial of the 8th order (Model 2) as well as with a harmonic function (Model 3) revealed model fit indices similar to the GAM (Model 4). All three models allowed for the two peaks in the probabilities for NAB (Fig. 2b). In contrast, the two mixed models revealed overall the best model fits. Due to their random intercept, which allowed for intra-individual variability in infant NAB, predicted mean values were lower by approximately two percentage points, albeit temporal patterns paralleling those of the GLMs and the GAM (Fig. 2b). Similar to infant sleep, GAMM (Model 6) revealed a slightly better fit than the GLMM with an 8th order polynomial and a random intercept (Model 5). Based on the GAMM, the probability of NAB started at 8 am with low values around 9% and reached then a first peak at noon with values

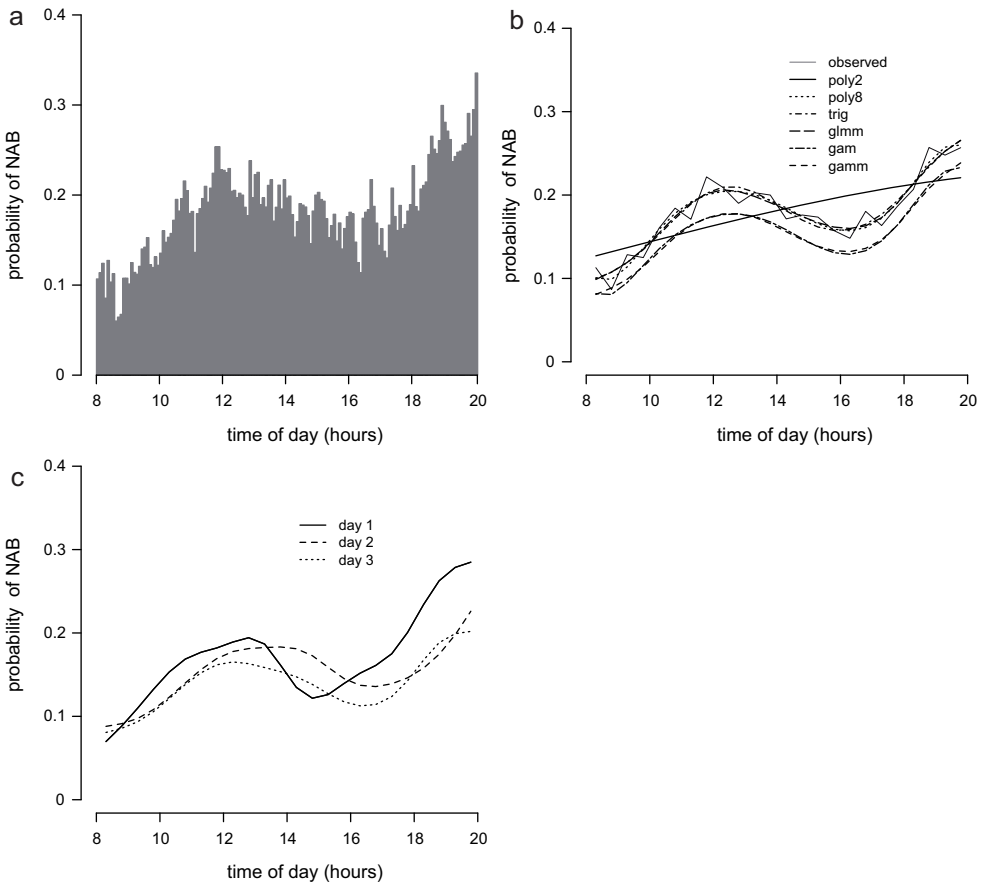

Fig. 2. (a) Spine plot showing the daily course of observed probabilities of infant NAB. (b) Daily course of predicted values of different models for probability of infant NAB. (c) Daily course of predicted values for probability of infant NAB for three consecutive days.

Table 4
Model fits for the probability of NAB.

Model	logLik ($\times 10^4$)	BIC ($\times 10^4$)	Parameters	Deviance explained (%)
0 GLM null model	−1.3397	2.6804	1	
1 GLM polynomial 2	−1.3312	2.6655	3	0.6
2 GLM polynomial 8	−1.3230	2.6553	9	1.2
3 GLM trigonometric	−1.3233	2.6538	7	1.2
4 GAM	−1.3243	2.6517	3	1.1
5 GLMM	−1.2128	2.4359	10	9.5
6 GAMM	−1.2141	2.4323	4	9.4

GLM null model = generalized linear model with intercept only; GLM polynomial 2 = GLM with a linear and quadratic polynomial; GLM polynomial 8 = GLM with a polynomial of the 8th and lower order; GLM trigonometric = GLM with a harmonic regression function; GAM = generalized additive model; GLMM = generalized linear mixed model with a polynomial of 8th and lower order and a random intercept; GAMM = generalized additive mixed model with a random intercept; logLik = log-likelihood of model; BIC = Bayesian information criterion; parameters = number of estimated parameters; deviance explained (%) = proportional change in deviance of a particular model relative to null model in percent.

around 19%. At 4 pm probabilities of NAB decreased (predicted mean values of 14%) and finally values reached the daily maximum of 26% at 8 pm, i.e., at sleep onset when data availability sharply dropped (preventing meaningful estimations past this time point).

Daily fluctuations in the probability of infant NAB were found among the three days as shown by significant main and interaction effects of day and time (Table 5 and Fig. 2c). Despite the found negative trend across the three days, the daily pattern for NAB remained similar across the study period and both the midday and evening peak, though different in temporal onset and magnitude, were found in each day. Approximately two percentage points difference were found for the midday peak during the three days and an average decrease of seven percentage points for the probability for NAB from the first to the third day.

4. Discussion

This is the first study to model temporal patterns of infant sleep or NAB using a variety of parametric and semi-parametric procedures simultaneously. Using high sensitive diary data in terms of temporal resolution and a mixed modeling approach we found similar behavior patterns in infant sleep and NAB as indicated by other researchers (Bamford et al., 1990; Barr, 1990; St. James-Roberts & Halil, 1991). Beside the well-known evening clustering in infant negative emotionality (Barr, 1990; St. James-Roberts & Halil, 1991) we also found a significant midday peak in infant NAB that was pervasive and only slightly affected by the study day. Overall, our results suggest that both types of mixed models, GAMM and GLMM with higher order polynomials, are particular powerful tools for estimating trajectories of infant activities such as sleep/wake behavior and NAB while allowing for the detection of variations within these behaviors in healthy infants.

4.1. Infant behavior modeling

To our knowledge, no study has reported 24-h patterns of sleep/wake behavior and NAB based on the semi-parametric procedure GAMM. The major advantages of the GAMM are its flexibility in dealing with non-linear patterns in infant behavior, its allowance for serially dependent data points by including random effects, and its smoothed transitions between study days in variables with a circadian rhythm. Our analyses indicate that GAMM fits the sleep and NAB data better than other parametric models that have previously been used for analyzing infant behavior. The higher polynomials of GLM and GLMM, although being able to fit almost any shape of a complex temporal pattern and generally being a flexible approach (Goldstein, 1979; Hoeksma & Koomen, 1992), produce model fits that are prone to overfitting and subsequently biased estimations of their individual parameters for which mostly no reasonable meaning can be attributed. I.e., due to the inclusion of too many parameters power is reduced to detect important predictors and complicated relations with interactions or nonlinear effects

Table 5
Model fits for probability of NAB: main and interaction effects for time of day and weekday using a GAMM.

	logLik ($\times 10^4$)	BIC ($\times 10^4$)	Parameters	Deviance change	Change in parameters	*p*-Value
NAB						
GAMM (random intercept) Intercept + time	−1.2141	2.4323	4			
GAMM (random intercept) Intercept + time + day	−1.2124	2.4310	6	34.1	2	<0.001
GAMM (random intercept) Inter-cept + time + day + time*day	−1.2113	2.4329	10	22.6	4	<0.001

GAMM = generalized additive model with a random intercept; parameter = number of estimated parameters; logLik = log-likelihood of model; deviance change = difference in $-2 \times$ logLik between model and previous model; change in parameters = difference in number of estimated parameters between model and previous model; BIC = Bayesian information criterion; time = 24 h period; day = study day.

between predictors and outcome variables may be prevailing that indeed exist in the sample, but not in the population (Babyak, 2004). Although this problem was not evident in our data, this could become a serious issue when additional covariates are included in the model. Unlike parametric models, GAMMs foster exploratory work that follows the principle of parsimony (Hawkins, 2004): on the one hand, they are complex enough to predict non-linear temporal patterns in infant behavior while discovering even small variations in individual trajectories by smoothing incidental variability in infant behavior patterns. On the other hand, they provide a reasonable interpretation of its model fits due to a low number of degrees of freedom, thereby bypassing the problem of overfitting. Consequently, the patterns found in the temporal data may be due to real variations rather than a reflection of an artifact, which makes them a powerful statistical tool (Wood, 2006). Furthermore, the inability of GLM and GLMM to smoothly connect the end and onset of different days in circadian patterns made them a less than perfect tool for modeling variability in variables with a circadian rhythm. Using cyclic cubic regression splines, GAMs and GAMMs are generally applicable for variables with a circadian rhythm. In GLMs and GAMs only, the violation of the underlying assumption of independence of observation and their resulting inability to model individual trajectories (Zuur et al., 2009) also made them inappropriate for modeling intra-individual variability in infant behavior. Despite the mentioned disadvantage, GLMs with a harmonic function are usually adequate for variables with a circadian rhythm and show an overall good model fit. Hence, if the functional relationship between predictors and outcomes is known and the variability among subjects is low, GLMs with a harmonic function are generally applicable. However, for variables like infant NAB, where no such functional relationship can be set up beforehand, GAMMs are the preferred procedure due to their allowance of non-linear and non-circadian trajectories and the inclusion of a random intercept. Note that GAMs/GAMMs can further be enhanced and/or modified, for example, using different smoothing functions or in the case of GAMMs by including additional random coefficients or specific correlation structures that define the serial dependence of the data.

4.2. Temporal patterns of infant sleep and NAB

We found similar patterns for sleep as published by other researcher for a sample of infants who were on average 5 months old (Bamford et al., 1990). The found pattern suggests consolidation of a circadian rhythm with a clear distinction between nighttime and daytime sleep (Davis et al., 2004; Iglowstein et al., 2003). The small peak in daytime sleep and the increased concentration of sleep during nighttime is characteristic for older infants (Iglowstein et al., 2003). NAB patterns followed a non-linear course with two significant peaks, the well-documented evening peak (Barr, 1990; St. James-Roberts & Halil, 1991) and an earlier peak occurring during midday, which all but the quadratic model predicted. The evening peak has been reported to be associated with developmental and environmental factors (Brazelton, 1962; St. James-Roberts & Halil, 1991). As the evening clustering in crying is found in western as well as in non-western societies (Barr, 1990; Barr, Konner, Bakeman, & Adamson, 1991; Hunziker & Barr, 1986) it may be assumed that the midday peak may also be an ubiquitous phenomenon in infants and have the same purpose as evening clustering of NAB. To our knowledge, only two other studies reported elevated crying amounts in their observed data before the evening, but they occurred during morning time (Brazelton, 1962; Rebelsky & Black, 1972). Both the midday and evening peak may be associated with overfatigue due to the natural sleep/wake pattern at these particular time points and consequently overload of sensation associated with immature self-soothing capabilities (DeSantis, Coster, Bigsby, & Lester, 2004). This idea is supported by the fact that both peaks occurred shortly before the probabilities of sleep increased. Infants of the ages included in our sample generally show an increased need for daytime sleep in the afternoon (Iglowstein et al., 2003). The smaller magnitude of the midday compared to the evening NAB peak may reflect less fatigue or sensory stimulation during the morning than during the rest of the day. In addition, the short duration of the daytime sleep may lack the same restorative effects that nighttime sleep possesses (Davis et al., 2004). Finally, Brazelton (1962) argued that the evening crying peak is associated with the peak period of tension in the family and the crying occurs due to the revoked attention by the naturally fatigued mother in the evening. Similarly, during midday the mother may be engaged in preparing lunch for the family and thus the attention is only partially focused on the infant, which may cause distress in the infant.

4.3. Daily fluctuations in infant sleep and NAB

Differences were found in the probability of infant sleep and NAB across the three study days. These findings are consistent with reported day-to-day variability in early infant behavior in various studies and seem to be most distinct in emotional expression (de Weerth & van Geert, 2002; Fish et al., 1991; St. James-Roberts & Halil, 1991; St. James-Roberts & Plewis, 1996; Wake, Morton-Allen, Poulakis, Hiscock, Gallagher, & Oberklaid, 2006). Day-to-day differences were rather small for infant sleep behavior indicating that the significance of these effects may be largely due to the statistical power of the models that are based on large amounts of data, rather than reflecting practically variations across the three days. Effects of daily fluctuations in NAB, however, were consistent and may be of theoretical importance because they may be a product of developmental, environmental, and interactional processes (de Weerth & van Geert, 2002; Fish et al., 1991; St. James-Roberts & Plewis, 1996). The found pattern of decrease of NAB over days points to reactivity effects based on the mothers and/or infants. First, mothers may have filled in the diary more conscientiously on the first day than on the second or third day although the amount and pattern of NAB was consistent across the three days. Second, because of the novelty of the task mothers may have been more focused on the completion of the diary rather than to attend and interact with their babies on the first day, which may have caused distress in the infant. Similarly, integrating the diary task with the manifold tasks of

childcare may have put additional stress on the mothers on the first day, which may have negatively affected the emotional state of the infant.

4.4. Limitations

Despite the apparent advantages of additive models, they are less efficient if the observed pattern is simple enough to allow for a parametric model or in cases where the kind of model is known beforehand. Furthermore, graphic display is highly important since the additive models do not have a predefined formulaic way to describe the relationship between predictor and outcome. However, if little is known about a relationship and there is no a priori reason for using a particular functional model, choosing a parametric model can easily lead to biased results due to overfitting of data or neglect of serially connected data points. Finally, the amount of crying was very low in our sample and crying less than 5 min was not recorded, which may be more prevalent in older infants. Hence, observed NAB patterns were predominantly based on infant fussing.

5. Conclusion

Our findings suggest unique behavior patterns in averaged five-months old infants. Beside the well-documented evening peak in NAB, we also found a midday peak. The semi-parametric model GAMM was thereby particularly useful to model these non-linear daily infant activities using a flexible approach, allowing for serial dependence of data points and for smoothing of transitions between the study days for circadian behaviors. Therefore, these kinds of models might be used as a standard procedure in analyzing infant data collected by measurement methods that produce a large amount of repeated-measure data points. For future research, GAMMs also allow the inclusion of predictors, such as personality or clinical variables that may explain between-individual differences in within-individual behavior trajectories.

Competing interests

None.

Acknowledgment

This work is part of the National Centre of Competence in Research (NCCR) Swiss Etiological Study of Adjustment and Mental Health (sesam). The Swiss National Science Foundation (SNF) (project no. 51A240-104890), the University of Basel, the F. Hoffmann-La Roche Corp. and Basel Scientific Society provided core support for the NCCR sesam. Special thanks to all families that have participated in the study and kindergartens, day nurseries, midwifes, and pediatricians who supported our study.

References

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*, 411–421.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade, & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Bamford, F. N., Bannister, R. P., Benjamin, C. M., Hillier, V. F., Ward, B. S., & Moore, W. M. (1990). Sleep in the first year of life. *Developmental Medicine and Child Neurology*, *32*(8), 718–724.

Barr, R. G. (1990). The normal crying curve: What do we really know? *Developmental Medicine and Child Neurology*, *32*(4), 356–362.

Barr, R. G., Konner, M., Bakeman, R., & Adamson, L. (1991). Crying in !Kung San infants: A test of the cultural specificity hypothesis. *Developmental Medicine and Child Neurology*, *33*(7), 601–610.

Barr, R. G., Kramer, M. S., Boisjoly, C., McVey-White, L., & Pless, I. B. (1988). Parental diary of infant cry and fuss behaviour. *Archives of Disease in Childhood*, *63*(4), 380–387.

Barr, R. G., Kramer, M. S., Pless, I. B., Boisjoly, C., & Leduc, D. (1989). Feeding and temperament as determinants of early infant crying/fussing behavior. *Pediatrics*, *84*(3), 514–521.

Bates, D. & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0. 999375-31. Available from: http://CRAN.R-project.org/package=lme4.

Belsky, J., Fish, M., & Isabella, R. (1991). Continuity and discontinuity in infant negative and positive emotionality: Family antecedents and attachment consequences. *Developmental Psychology*, *27*(3), 421–431.

Brazelton, T. B. (1962). Crying in infancy. *Pediatrics*, *29*, 579–588.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*(3), 371–403.

Davis, K. F., Parker, K. P., & Montgomery, G. L. (2004). Sleep in infants and young children: Part one: Normal sleep. *Journal of Pediatric Health Care*, *18*(2), 65–71.

de Weerth, C., & van Geert, P. (2002). Changing patterns of infant behavior and mother-infant interaction: Intra- and interindividual variability. *Infant Behavior and Development*, *24*, 347–371.

DeSantis, A., Coster, W., Bigsby, R., & Lester, B. (2004). Colic and fussing in infancy, and sensory processing at 3 to 8 years of age. *Infant Mental Health Journal*, *25*, 522–539.

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, *156*(3), 193–203.

Emde, R. N., Gaensbauer, T. J., & Harmon, A. (1976). *Emotional expression in infancy: A biobehavioral study*. New York: International University Press.

Faraway, J. J. (2006). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall.

Fish, M., Stifter, C. A., & Belsky, J. (1991). Conditions of continuity and discontinuity in infant negative emotionality: Newborn to five months. *Child Development*, *62*(6), 125–1537.

Forsyth, B. W., Leventhal, J. M., & McCarthy, P. L. (1985). Mothers' perceptions of problems of feeding and crying behaviors. *American Journal of Diseases of Children*, *139*, 269–272.
Goldstein, H. (1979). *The design and analysis of longitudinal studies*. London: Academic Press.
Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distribution: Setting the scene. *Ecological Modelling*, *157*, 89–100.
Hastie, T. J., & Tibshirani, R. J. (1986). Generalized additive models. *Statistical Science*, *1*, 297–318.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall.
Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Science*, *44*, 11–12.
Hoeksma, J. B., & Koomen, H. M. Y. (1992). Multilevel models in developmental psychological research: Rationales and applications. *Infant and Child Development*, *1*(3), 157–167.
Holditch-Davis, D., Edwards, L. J., & Helms, R. W. (1998). Modeling development of sleep-wake behaviors: I. Using the mixed general linear model. *Physiological Behavior*, *63*(3), 311–318.
Hunziker, U. A., & Barr, R. G. (1986). Increased carrying reduces infant crying: A randomized controlled trial. *Pediatrics*, *77*(5), 641–648.
Iglowstein, I., Jenni, O. G., Molinari, L., & Largo, R. H. (2003). Sleep duration from infancy to adolescence: Reference values and generational trends. *Pediatrics*, *111*(2), 302–307.
Lam, J., Barr, R. G., Catherine, N., Tsui, H., Hahnhaussen, C. L., Pauwels, J., et al. (2010). Electronic and paper diary recording of infant and caregiver behaviors. *Journal of Developmental Behavioral Pediatrics*, *31*(9), 685–693.
Littell, R. C., Miliken, G. A., Stroup, W. W., Wofinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC, USA: SAS Institute Inc.
Lummaa, V., Vuorisalo, T., Barr, R. G., & Lehtonen, L. (1998). Why cry? Adaptive significance of intensive crying in human infants. *Evolution and Human Behavior*, *19*, 193–202.
McCullagh, P., & Nelder, J. (1983). *Generalized linear models*. London: Chapman and Hall.
Miller, A. R., Barr, R. G., & Eaton, W. O. (1993). Crying and motor behavior of six-week-old infants and postpartum maternal mood. *Pediatrics*, *92*(4), 551–558.
Müller, S., Hemmi, M. H., Wilhelm, F. H., Barr, R. G., & Schneider, S. (2011). Parental report of infant sleep behavior by electronic versus paper-and-pencil diaries, and their relationship to actigraphic sleep measurement. *Journal of Sleep Research*,. Epub ahead of print
Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of Royal Statistical Society Series A*, *135*, 370–384.
R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing.
Rebelsky, F., & Black, R. (1972). Crying in infancy. *Journal of Genetic Psychology*, *121*(1), 49–57.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: University Press.
St. James-Roberts, I., Conroy, S., & Wilsher, C. (1998). Stability and outcome of persistent infant crying. *Infant Behavoir and Development*, *21*(3), 411–435.
St. James-Roberts, I., & Halil, T. (1991). Infant crying patterns in the first year: Normal community and clinical findings. *Journal of Child Psychology and Psychiatry*, *32*(6), 951–968.
St. James-Roberts, I., Hurry, J., & Bowyer, J. (1993). Objective confirmation of crying durations in infants referred for excessive crying. *Archives of Diseases of Children*, *68*(1), 82–84.
St. James-Roberts, I., & Plewis, I. (1996). Individual differences, daily fluctuations, and developmental changes in amounts of infant waking, fussing, crying, feeding, and sleeping. *Child Development*, *67*(5), 2527–2540.
van den Boom, D. C., & Hoeksma, J. B. (1994). The effect of infant irritability on mother-infant interaction: A growth-curve analysis. *Developmental Psychology*, *30*(4), 581–590.
van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development*, *5*, 340–374.
Wake, M., Morton-Allen, E., Poulakis, Z., Hiscock, H., Gallagher, S., & Oberklaid, F. (2006). Prevalence, stability, and outcomes of cry-fuss and sleep problems in the first 2 years of life: Prospective community-based study. *Pediatrics*, *117*(3), 836–842.
Wood, S. (2009). gamm4: Generalized additive mixed models using mgcv and lme4. R package version 0. 0-1. Available from: http://CRAN.R-project.org/package=gamm4.
Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.
Yee, T. W., & Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, *2*(58), 7–602.
Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.