Prediction of 3-year developmental outcomes from sleep development over the preterm period

Diane Holditch-Davis*, Michael Belyea, Lloyd J. Edwards

CB# 7460 Carrington Hall, School of Nursing and Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7460, USA

Received 15 August 2004; received in revised form 5 December 2004; accepted 10 December 2004

Abstract

This study explored whether the developmental status of 3-year-old prematurely born children could be predicted from the development of sleep–wake behaviors during the preterm period. Sleep–wake states were observed weekly in 71 preterms from the time they were no longer critically ill until term or discharge, and the general linear mixed model was used to model the development of preterm behaviors for both individuals and the group. At 3 years, the home environment and developmental status of 51 children was assessed. Individual infants’ deviations from group slopes and intercepts of eight preterm sleep variables were used as predictors. Cluster analyses of the predictor variables were found to relate to Stanford-Binet IQ, language, and fine and gross motor abilities at 3 years post-term. Children in Cluster 4, which showed more rapid active sleep development, averaged higher IQs and better language and fine motor abilities than children in other clusters.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Developmental outcomes; Prediction; Sleep development; Respiration; Premature infants; Biological risk

1. Introduction

Nearly 50% of premature infants who either require mechanical ventilation or weight less than 1500 g have at least mild developmental problems by school age, including learning disabilities or behavioral problems (Bhutta, Cleves, Casey, Cradock, & Anand, 2002; Dezoete, MacArthur, & Tuck, 2003; O’Brien...
et al., 2004). Yet many prematures develop normally, and most experience only minor problems. Prediction of outcome for individuals, except in the case of extreme handicap, is not yet feasible.

Many investigators have found that the social environment is a better predictor of infant outcome than perinatal complications (Gross, Mettelman, Dye, & Slagle, 2001; Hack et al., 1992). This inability to make predictions from biological risk indices probably occurs because infant outcome results from interactions among multiple factors, including genetic background, medical severity, and social environment. Fetal and neonatal brain development is an “experience expectant” process in which species-typical experiences enable the brain to make structural and functional changes needed for the next developmental stages (Greenough, Black, & Wallace, 1987). To maintain a balance between the needs of the present stage and needs of subsequent stages, this process is somewhat plastic (Oppenheim, 1981). Placing a young infant in the atypical environment of the neonatal intensive care unit may initially result in adaptation but lead to changes that are maladaptive at older ages. A more fruitful approach to predicting outcomes may be to identify indices of neural integrity.

The development of sleeping in preterm infants has the potential to predict long-term outcomes. Sleep–wake development in the preterm period—decreases in the amount of active sleep, increases in quiet sleep and waking states, and increases in sleep state organization—parallels brain development (Blumberg & Lucas, 1996; Curzi-Dascalova, Peirano, & Morel-Kahn, 1988; Ingersoll & Thoman, 1999; Sahni, Schulze, Stefanski, Myers, & Fifer, 1995). The state patterns of preterms with neurological insults or other medical complications differ from those of healthier infants (Doussard-Roosevelt, Porges, & McClenney, 1996).

Sleeping and waking have also been used to predict cognitive and neurological outcomes. Sleep–wake measures during the preterm period—including the amount of crying, sleep cycle length, state organization quality, and amount of night sleep—predict Bayley scores during the first year (Borghese, Minard, & Thoman, 1995; DiPietro & Porges, 1991; Fajardo, Browning, Fisher, & Paton, 1992; Gertner et al., 2002). Developmental changes in the amounts of specific sleep behaviors during the first year are related to outcomes in the second year (Anders, Keener, & Kraemer, 1985; Whitney & Thoman, 1993). Further, the stability of sleep–wake patterns in the neonatal period predicts later development (Thoman, Denenberg, Sieval, Zeidner, & Becker, 1981; Whitney & Thoman, 1993). EEG sleep measures at term and in healthy preterms are related to developmental outcome at up to 8 years (Cohen, Parmelee, Beckwith, & Sigman, 1986; Ferrari et al., 1992; Scher, Steppe, & Banks, 1994).

These findings indicate that sleep behaviors are potentially useful for examining brain function in relation to later development. However, to date associations between preterm sleep patterns and long-term development have been small, partly because most studies used measures from a single age. Indices that examine changes over age in sleep behaviors might be more effective since they incorporate both the current functioning of the central nervous system (CNS) and the degree to which the CNS exhibits normal development despite insults (Thoman, 1982).

The purpose of this study was to explore whether the developmental status of 3-year-old prematurely born children could be predicted from the development of sleep behaviors during the preterm period. Development at 3 years was assessed along multiple dimensions: language, cognition, and fine and gross motor skills. The social environment of the children was controlled for in all predictions. The deviations of individuals from the intercepts and slopes of group developmental patterns during the preterm period for eight sleep behaviors were the predictors. These deviations indicate the degree to which infants are able to show typical development, as compared to other preterms, despite whatever neurological insults they may have experienced.
2. Methods

2.1. Participants

The participants were 71 preterm infants who either weighing less than 1500 g at birth or requiring mechanical ventilation or continuous positive airway pressure; 36 had both problems. They were born between October 1985 and March 1990 and hospitalized in a regional perinatal center in the southeast. Infants with congenital problems associated with neurological problems (e.g., Down Syndrome or microcephaly) were excluded. Demographic characteristics of these infants have been previously reported (Holditch-Davis & Edwards, 1998).

Sixty infants were followed at 3 years. Of the infants not followed, three had died, two could not be located, and parents of the rest refused follow-up. The 51 children who had an assessment of the social environment at 3 years and at least two observations during the preterm period are the focus of this report. These children had 47 primary caretakers—43 mothers and 4 grandmothers. Three grandmothers had legal custody; the fourth was the primary caretaker for a child who was in the father’s custody after a divorce. The infants averaged 28.2 weeks gestational age (S.D. 2.1) and a birthweight of 1106 g (S.D. 319); 86.3% were average size for gestational age. They were 54.9% male and were 47.1% African-American, 49.0% white, and 5.9% American-Indian. They averaged 10.2 days of mechanical ventilation (S.D. 12.6). During the neonatal hospitalization, 39.2% had an intraventricular hemorrhage (23.5% Grade I, 3.9% Grade II, 7.8% Grade III, and 3.9% Grade IV), 3.9% had periventricular leukomalacia, 33.3% had chronic lung disease, 29.4% had major surgery, and 86.3% were treated with theophylline. Four sets twins were in the study, and seven additional twins and one triplet were in the study without their siblings; 54.9% were firstborn children. Their caretakers averaged 30.6 years (S.D. 7.2), and 70% were married. They averaged 13.0 years of education (S.D. 2.2).

2.2. Infant characteristics

The mothers completed a questionnaire on family demographic characteristics at enrollment and again at 3 years. Other infant characteristics and medical information was determined from the neonatal medical record. Gestational age at birth was calculated from the obstetric estimated date of confinement, determined from either the date of the mother’s last menstrual period or an ultrasound examination. If this gestational age did not agree within 2 weeks with a gestational age assessment (Ballard et al., 1991) conducted by a pediatrician on admission to the intensive care unit, the age from the assessment was used. Post-conceptional age was the number of weeks since birth added to gestational age.

2.3. Preterm sleep behaviors

The variables used to predict 3-year outcomes were derived from the results of analyses of developmental patterns of preterm behaviors conducted in previous publications (Holditch-Davis, Brandon, & Schwartz, 2003; Holditch-Davis & Edwards, 1998; Holditch-Davis, Edwards, & Wigger, 1994). Briefly, the two sleep states, active sleep and quiet sleep, have been shown to exhibit reliable individual differences for preterms (Holditch-Davis & Edwards, 1998). The sleep states were judged based on eye opening, general body movements, muscle tone, eye movement, and respiration (Holditch-Davis & Edwards, 1998). One critical criteria of each sleep state was used to indicate organization. For active sleep, the presence
of rapid eye movements was scored every 10 s. Active sleep without REM was considered to be poorly organized active sleep. Respiration regularity indicated the organization of quiet sleep. The respiration tape was scored visually for respiration regularity during quiet sleep using a three-point scale of very regular respiration, regular respiration, and irregular respiration (Holditch-Davis & Edwards, 1998). Two infant behaviors were also studied: sigh (a deep audible respiration of the infant) and negative facial expression (cry face or frown, occurring in isolation or with crying; Holditch-Davis et al., 2003).

To account for slight variations in observation length, sleep states and infant behaviors were measured as percentages of the 4-h observation. The amount of active sleep without REM was measured as a percent of active sleep time. Respiration regularity variables were measured as percents of quiet sleep. These variables had similar developmental patterns: the regular respiration variables increased, and irregular respiration decreased (Holditch-Davis, 1990). Therefore, they were combined by adding two times the percent of very regular respiration plus the percent of regular respiration minus the percent of irregular respiration. Possible scores on quiet sleep respiration regularity ranged from −100 to 200.

The final two predictors were the average length in seconds of respiratory pauses in each sleep state. A respiratory pause was defined as: absence of respiratory activity for more than 2 but less than 20 s between the end of expiration and the beginning of inspiration and obvious disruption of the ongoing respiratory pattern without associated hypoxia (Holditch-Davis et al., 1994). Because respiratory pauses were identified by a lack of breathing movements, only central apnea was measured as a respiratory pause. Respiratory pause length was measured from the end of expiration to the beginning of inspiration.

2.4. Three-year instruments

2.4.1. Home

The Home Observation for Measurement of the Environment (HOME, 0–3 year version) is designed to identify young children at risk for developmental delay because their environment fails to provide the stimulation needed for learning (Caldwell & Bradley, 1980, 1984). The 45 binary items are scored using a combination of semi-structured mother interview about the child’s routine activities, observation of mother and child interaction, and assessment of kinds of play materials available to the child. Scores for the HOME range from 0 to 45, with higher scores indicating a more stimulating environment. The HOME total score showed internal consistency of .88 in this study (Tesh & Holditch-Davis, 1997).

2.4.2. IQ

The Stanford-Binet Intelligence Scale (4th edition) is a continuous scale for the scoring of cognitive status (Thorndike, Hagan, & Sattler, 1986). It has a mean of 100 and a standard deviation of 16. IQs of 83 and below indicate low cognitive abilities; IQs of 84 and above are within the normal range. Also, subscale scores are obtained in verbal reasoning, abstract reasoning, quantitative reasoning, and memory. IQs of four children in our sample were estimated from scores on subscales because these children could not be tested on all items. Three children who were untestable because of severe cerebral palsy and major developmental delays were given an IQ of 50 (Table 1).

2.4.3. PEET

The Pediatric Extended Examination at Three (PEET) was designed for early detection of learning, attention, and behavioral problems in 3- to 4-year-old children (Blackman, Levine, Markowitz, & Aufseeser, 1983). The tool uses developmental observation and the physical examination to assess status in
Table 1
Mean scores (and standard deviations and ranges) of the sample of 3-year old prematurely born children on a variety of developmental tests

<table>
<thead>
<tr>
<th>Instrument</th>
<th>N</th>
<th>Mean (S.D.)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford-Binet IQ</td>
<td>51</td>
<td>86.5 (16.3)</td>
<td>50–117</td>
</tr>
<tr>
<td>PEET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive language</td>
<td>45</td>
<td>11.5 (3.6)</td>
<td>2–17</td>
</tr>
<tr>
<td>Reception</td>
<td>45</td>
<td>24.7 (8.4)</td>
<td>11–45</td>
</tr>
<tr>
<td>Fine motor output</td>
<td>48</td>
<td>4.5 (3.2)</td>
<td>0–14</td>
</tr>
<tr>
<td>Gross motor output</td>
<td>48</td>
<td>10.1 (4.5)</td>
<td>0–17</td>
</tr>
<tr>
<td>Reactions to assessment</td>
<td>37</td>
<td>42.2 (6.6)</td>
<td>32–54</td>
</tr>
<tr>
<td>ANSER attention questionnaire</td>
<td>46</td>
<td>8.9 (7.3)</td>
<td>0–30</td>
</tr>
<tr>
<td>CBCL total</td>
<td>46</td>
<td>37.7 (20.2)</td>
<td>8–99</td>
</tr>
<tr>
<td>HOME</td>
<td>51</td>
<td>36.6 (6.8)</td>
<td>17–45</td>
</tr>
</tbody>
</table>

In addition, the child’s attention is scored on a 19-item, Reactions During Assessment scale. Each item in the skill domains was scored on a four-point scale (falling below levels, or at Levels 1, 2, or 3). Level 2 is normal behavior for 3 and half-year-old children. For scoring, we used the item analysis data in which each item receives a score between 0 (below levels) and 3 (level 3), and these values summed. Possible scores were 0–48 in reception, 0–18 in fine motor output, 0–15 in gross motor output, and 0–19 in expressive language. Children refusing more than 1 item in a domain were not scored.

In a subset of 21 children, the raw motor score on the McCarthy Scales of Children’s Abilities (McCarthy, 1971) was correlated with the fine motor score ($r = .74, p < .001$). Gross motor score ($r = .30$) was not related to the McCarthy motor scores. Expressive language score was related to reception ($r(52) = .91, p < .01$), verbal reasoning on the Stanford-Binet ($r(50) = .78, p < .001$), and nurse and psychologist examiners’ clinical judgments of language concerns ($-.33$ and $-.53$, respectively, $n = 52, p < .05$). Reactions was related to verbal reasoning on the Stanford-Binet ($r(50) = .75, p < .001$) and the examiners’ judgments of language concerns ($-.33$ and $-.53$, respectively, $n = 52, p < .05$). Due to similarity between the two language scores, only the expressive language score was used in analyses.

The Reactions to Assessment Scale scores range from 19 to 57 with higher scores indicating better attention. Scores of 36 or below indicate concerns in the area of attention. Internal consistency for the Reactions to Assessment scale was .88.

2.4.4. CBCL
The Child Behavior Checklist for Ages 2–3 (CBCL) is a parent-completed instrument designed to detect behavior problems in 2- and 3-year-old children (Achenbach, Edelbrock, & Howell, 1987). The checklist results in a total behavior problem score and several sub-scale scores. Scores above the 98th percentile are considered to be clinically significant.

2.4.5. Attention questionnaire
The Aggregate Neurobehavioral Student Health and Educational Review (ANSER System) is a group of parental data collection instruments for preschool through school-aged children (Levine, 1981).
In this study, we used 20 questions about attention problems and hyperactivity from the ANSER System. A 21st question about underactive behavior was eliminated since it did not relate to the rest of the scale, even when reverse scored. Each question is scored on a three-point scale: 0 does not apply, 1 applies somewhat, 2 definitely applies. The overall score ranges from 0 to 40. Internal consistency for the attention questionnaire was .90.

2.5. Procedures

The study was approved by the institutional committee for protection of human subjects. Infants were enrolled when their medical conditions were no longer critical (not on mechanical ventilation or in an immediate life-threatening situation) if an additional hospital stay of at least 1 week was anticipated and consent was obtained from the parents. Mothers completed the demographic questionnaire after enrollment.

Infant sleep and behaviors were observed once a week from about 19:00 to 23:00 h until hospital discharge or term age. During the observations, the occurrences of sleep states and other behaviors were recorded every 10 s. The end of each 10-s period was signaled audibly through an earphone from a small electronic timer. At this signal, the predominant sleep–wake state for the period and any behaviors that occurred were recorded into the event recorder. Multiple occurrences of the same behavior in the same epoch were not recorded. Each observation was conducted by one of two observers. Also, the infant’s respiration was recorded on a Gulton chart recorder with a piezo electric sensor pad placed under the infant’s crib pad, in order to identify apnea and respiration regularity. Nothing was attached to the infant, and the infant continued on normal heart and apnea monitors.

At 3 years post-term, the HOME Inventory was scored during a home visit (Tesh & Holditch-Davis, 1997). When both children from a twin pair were in the study, the HOME was scored on each of them at different times. One month later, a developmental assessment of the child was conducted at a developmental evaluation center. (Two children were tested in their homes: one had moved out of state and the other repeatedly missed appointments at the center.) The Stanford-Binet Intelligence Test and PEET were administered by two different examiners, neither of whom was aware of the child’s history. The mother also completed the demographic questionnaire, Child Behavior Checklist, and ANSER questionnaire.

2.6. Data analysis

In previous papers (Holditch-Davis & Edwards, 1998), the general linear mixed model, or hierarchical linear model, was used to determine the developmental patterns of the preterm variables by regressing the amounts of these variables over post-conceptional age. The mixed model has three components: the fixed effects component, random effects component, and random error component (Edwards, Stewart, Muller, & Helms, 2001; Holditch-Davis, Edwards, & Helms, 1998). The fixed effects and random error components are analogous to the corresponding components of a standard multiple regression. The fixed effects component represents a population regression line with respect to post-conceptional age. The random effects component for each subject is the difference between the subject’s regression and the population regression and is a measure of how this subject differs from the population regression. The mixed model easily accommodates data that is collected at different ages and different lengths of time for each subject.
Previous studies found that most preterm variables showed changes over age (Holditch-Davis & Edwards, 1998). Eight variables with significant development over the preterm period, with slopes that did not differ between two halves of the sample, and that accounted for reasonable amounts of variance were chosen as predictors for 3-year outcomes. Active sleep decreased over the preterm period from a mean of about 82% at 27 weeks to 62% at 39 weeks, whereas quiet sleep increased from 13 to 23% (Holditch-Davis & Edwards, 1998). Sighs decreased over the preterm period; negative facial expressions increased (Holditch-Davis et al., 2003). Active sleep without REM decreased from a mean of 58% at 27 weeks to 42% at 39 weeks (Holditch-Davis & Edwards, 1998). Mean quiet sleep respiration regularity scores increased from 30 at 27 weeks to 90 at 39 weeks (Holditch-Davis & Edwards, 1998). Respiratory pause length decreased in active sleep from a mean of 3.8 s at 27 weeks to 3.0 s at 39 weeks and in quiet sleep from 5.2 to 3.6 s (Holditch-Davis et al., 1994).

Individual infants’ deviations from the group on slope and intercept (random effects) from the mixed models were used as predictors. The amounts of five variables—active sleep without REM, quiet sleep regularity score, mean length of respiratory pauses in quiet sleep, sigh, and negative facial expressions—were affected by one or two covariates, such as mechanical ventilation or gender, but no covariate affected any variable’s developmental pattern (Holditch-Davis & Edwards, 1998). Mixed model deviations take into account these significant covariates in the model. Thus, each participant’s deviations are from a group with similar covariate characteristics. Only one variable—negative facial expressions—had a large correlation between its random intercept and random slope, .81. For this variable, only the random intercept effect was used as a predictor. Thus, the 15 predictor variables were the random slope and intercept for active sleep, quiet sleep, active sleep without REM, quiet sleep regularity score, mean length of respiratory pauses in active sleep, mean length of respiratory pauses in quiet sleep, and sigh and the random intercept for negative facial expressions.

To make useful predictions, preterm variables need to be combined in a way that would identify at-risk individuals. Therefore, we used a profile-oriented approach to identify infants that shared common preterm developmental patterns and, thus, similar biological risk (Magnusson, 1995). A hierarchical cluster analysis, using Ward’s method, was conducted using the preterm deviations. Ward’s method was selected because it optimizes the minimum variance within clusters (Aldenderfer & Blashfield, 1984). Each variable was standardized to have a mean of zero and standard deviation of one to eliminate scaling effects. A stepwise discriminant analysis was applied to the cluster solution to determine the number of significant clusters and which variables established each cluster.

3. Results

3.1. Three-year developmental outcomes

Table 1 presents the outcomes of the 51 children followed at 3 years. They had a high rate of developmental problems, but most problems were mild. Only five children had IQs in the mentally retarded range (below 70), with four of them scoring below 50. Three of these children had severe cerebral palsy. The fourth came from an extremely unsupportive home situation and was hyperactive. He probably could not be adequately tested, and the score below 50 was an underestimation of his abilities. There was a general lowering of expected IQs for all of the children. Although several of the parents were professionals, the highest recorded IQ was 117, and the group mean was almost 1 S.D. below the normative mean.
Preliminary analyses indicated that three outcome variables were significantly correlated with the HOME score: IQ, expressive language, and the CBCL. The HOME was also possibly correlated with the PEET Reactions to Assessment and the fine motor output as $p < .10$.

3.2. Prediction of 3-year outcomes

A cluster analysis was conducted using the preterm deviations in slope and intercept. All 60 infants with some data at age 3 years, except the four infants with only one preterm observation, were used for this analysis since the focus was on preterm behaviors. The infants formed four clusters that had discriminant function eigenvalues of 1 or greater. Function 1 accounted for 48% of the variance and consisted primarily of the slope and intercept of quiet sleep and the slope of the quiet sleep regularity score and sigh. Function 2 accounted for 40% of the variance and consisted of slope and intercept of mean respiratory pause length in active sleep, the intercept of quiet sleep regularity, and the slope of mean respiratory pause length in quiet sleep. Function 3 accounted for 11% of the variance and was made up of the slope and intercept for active sleep organization, slope for active sleep, and intercept for negative facial expressions.

The 22 infants in Cluster 1 showed slower development of respiratory pauses and sighs. The seven infants in Cluster 2 showed more rapid quiet sleep organization and more rapid respiratory pause development. The 19 infants in Cluster 3 showed slower development of both active and quiet sleep organization, and the eight infants in Cluster 4 showed more rapid active sleep development and longer respiratory pauses.

To see if these clusters could have been identified by medical and demographic variables available in the preterm period, the clusters were compared on demographic and medical complications using linear regressions. The four clusters did not differ on birthweight, gestational age at birth, gender, race, presence or severity of intraventricular hemorrhage, PVL, length of mechanical ventilation, or incidence of chronic lung disease.

The three infants with IQs below 50 and severe cerebral palsy were in Cluster 3. The child with a questionable IQ below 50 was in Cluster 4. Two additional children with cerebral palsy in the sample but with near to normal IQs were in Cluster 2.

Finally, we compared the four clusters on 3-year outcomes using multiple regression analyses for each outcome with the HOME score as a covariate (see Table 2). The analysis excluded the three children with cerebral palsy and IQs below 50. The child whose IQ below 50 was questionable was excluded from only the IQ analysis. The children in Cluster 4 had the best outcomes. They had the highest mean IQ, which was significantly higher than that for either Clusters 1 or 3. The children in Cluster 4 also had significantly better scores on language and fine motor skills than the other three clusters. Although the overall group comparison was not significant for gross motor abilities, children in Cluster 4 had significantly higher gross motor scores than children in Clusters 1 and 2. The children in Cluster 2 had significantly higher examiner ratings of attention than Cluster 1.

3.3. Exploratory regression analyses

To determine how much of the variance in developmental outcomes could be explained by the preterm variables, we conducted exploratory regressions using all sleep–wake behavioral developmental variables to predict IQ, language, fine motor abilities, and gross motor abilities in the 51 children (see Table 3). We used a backward elimination process whereby variables with $p > .10$ were removed from a preliminary
Table 2

Results of multiple regression analyses comparing the children in the four clusters on 3-year developmental outcomes

<table>
<thead>
<tr>
<th>Domain</th>
<th>Results of multiple regression analyses</th>
<th>Overall</th>
<th>Cluster</th>
<th>Least square means</th>
<th>Post-hoc comparisons</th>
<th>d.f.</th>
<th>F</th>
<th>d.f.</th>
<th>F</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Significant differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td>4.42</td>
<td>5.22**</td>
<td>3.46</td>
<td>3.13**</td>
<td>89.4</td>
<td>92.0</td>
<td>85.3</td>
<td>102.2</td>
<td>4&gt;1, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td>4.40</td>
<td>4.36**</td>
<td>3.40</td>
<td>2.98</td>
<td>11.8</td>
<td>9.8</td>
<td>10.6</td>
<td>14.4</td>
<td>4&gt;2, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fine motor</td>
<td></td>
<td>4.40</td>
<td>4.38**</td>
<td>3.40</td>
<td>4.44**</td>
<td>4.0</td>
<td>4.3</td>
<td>4.7</td>
<td>8.5</td>
<td>4&gt;1, 2, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross motor</td>
<td></td>
<td>4.40</td>
<td>2.25**</td>
<td>3.40</td>
<td>2.72</td>
<td>9.6</td>
<td>9.7</td>
<td>12.0</td>
<td>13.4</td>
<td>4&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEET attention</td>
<td></td>
<td>4.32</td>
<td>2.19</td>
<td>3.32</td>
<td>1.65</td>
<td>41.3</td>
<td>51.7</td>
<td>43.1</td>
<td>41.5</td>
<td>2&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANSER attention</td>
<td></td>
<td>4.39</td>
<td>7.4</td>
<td>3.39</td>
<td>.43</td>
<td>8.1</td>
<td>10.5</td>
<td>10.7</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBCL</td>
<td></td>
<td>4.39</td>
<td>1.65</td>
<td>3.39</td>
<td>.97</td>
<td>33.1</td>
<td>41.1</td>
<td>44.7</td>
<td>37.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Each line in table is a separate multiple regression analysis. *p < .10, **p < .05, ***p < .01.

Table 3

Results of step-wise multiple regression analyses predicting on 3-year developmental outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>d.f.</th>
<th>Adjusted $R^2$</th>
<th>Adjusted $R^2$ w/out HOME</th>
<th>Intercept Parameter</th>
<th>HOME Parameter</th>
<th>Other Predictors Parameter</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>2.49</td>
<td>.26***</td>
<td>.09</td>
<td>52.76***</td>
<td>11.18 .91**</td>
<td>.30 Inter: Neg Fac Expr</td>
<td>2.72**</td>
</tr>
<tr>
<td>Expressive</td>
<td>4.41</td>
<td>.29**</td>
<td>.14</td>
<td>2.81</td>
<td>2.92 .23**</td>
<td>.08 Inter: AS Resp Pause</td>
<td>4.59**</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inter: No REM AS</td>
<td>.44**</td>
</tr>
<tr>
<td>Fine motor</td>
<td>8.37</td>
<td>.59***</td>
<td>.43***</td>
<td>– 1.95</td>
<td>1.98 .17**</td>
<td>.05 Interc Quiet Sleep</td>
<td>–1.03**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inter: QS Resp Reg</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interc Neg Fac Expr Sleep</td>
<td>–1.07***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Slope Quiet Sleep AS Resp</td>
<td>–4.04**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Slope AS Resp Pause</td>
<td>–33.51***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Slope QS Resp Reg</td>
<td>.72</td>
</tr>
</tbody>
</table>

Note: Interc, intercept; AS, active sleep; QS, quiet sleep; Resp Reg, respiration regularity score; Resp Pause, mean length of respiratory pauses; No REM AS, active sleep without REMs; Neg Fac Expr, negative facial expressions.

* p < .05.
** p < .01.
*** p < .001.
model; variables with \( p > .05 \) were removed from the final model. The 3-year HOME score was used as a covariate and remained in the equation regardless of probability level. The preterm variables and HOME significantly predicted IQ, expressive language and fine motor skills, but there were no significant predictors for gross motor abilities. When the variance due to the 3-year HOME score was excluded, the regressions for IQ, fine motor, and expressive language abilities remained significant, though only a small amount of variance in IQ and expressive language was explained.

4. Discussion

The development of sleep during the preterm period can be used to predict outcome 3-years later when the quality of the home environment is controlled. The development of sleep behaviors probably reflects the status of infant’s CNS (Halpern, MacLean, & Baumeister, 1995; Thoman, 1982). Our findings also indicated that several behaviors were related to outcomes: all eight predictor variables were used in developing clusters, and all but respiratory pause length in quiet sleep were related to at least one outcome in the multiple regressions. The clusters of the preterm sleep behaviors that indicated risk for development problems at 3 years could not be identified from preterm medical or demographic variables.

Moreover, the developmental pattern differentiating the cluster with the best development outcomes, more rapid development of active sleep, makes theoretical sense. The dramatic developmental decrease in active sleep over the preterm period makes possible the increases in quiet sleep and waking states (Holditch-Davis & Edwards, 1998; Ingersoll & Thoman, 1999). Several other researchers have also found that the active sleep or its criteria are related to developmental problems in preterm and fullterm infants (Becker & Thoman, 1981; Whitney & Thoman, 1993) although other researchers have found that other sleeping and waking behaviors—particularly waking, crying, or quiet sleep—were better predictors of developmental outcomes (Anders et al., 1985; DiPietro & Porges, 1991; Gertner et al., 2002). However, only Anders et al. and our study examined the developmental patterns of these behaviors, rather than the amounts at specific ages.

Although other researchers have used sleeping and waking to predict developmental outcome in pre-matures, ours is one of the few studies to show that outcomes can be predicted from sleep patterns in the preterm period to an age at which cognitive status can be reliably determined (McCall, 1983) and mild delays and hyperactivity can be identified (Astbury et al., 1985). Most studies have either predicted from the preterm period to sometime during the first year (DiPietro & Porges, 1991; Fajardo et al., 1992; Gertner et al., 2002) or from later in the first year to the preschool period (Whitney & Thoman, 1993). Yet prediction from the preterm period to the preschool period is important so that infants can be assessed at a time when they are available to health care professionals and so that early intervention can be targeted to those infants in greatest need.

In general, we found that outcomes with a strong neurological component, such as IQ, expressive language, and fine motor abilities, were more likely to be predicted by preterm sleep development than outcomes primarily influenced by social situations. Maternal perceptions of attention and behavioral problems did not differ among the preterm clusters. These outcomes probably reflect the child and mother together, rather than just the child’s CNS.

An exception was gross motor abilities, which has a strong neurological component but did not show a significant relationship with cluster membership or a significant regression analysis. Post-hoc analyses did show significant differences between clusters 4 and 1, suggesting that at least some aspects of gross
motor abilities at 3 were related to preterm sleep behaviors. Possibly, by 3 years, most children, except those with cerebral palsy, have achieved so many of the gross motor milestones that gross motor abilities do not vary significantly.

The techniques used to assess preterms must be applicable to infants of varying post-conceptional ages since the ages at which infants become healthy enough to study and the ages at which they are discharged from the hospital vary. Thus, there is no single age at which all, or even most, infants will be available for study. In our study, intercepts were as often related to outcomes as slopes. Since we used the intercept at 33 weeks, conducting a single observation at that age might have been simpler, but only 34 of the 71 infants in our larger study (Holditch-Davis & Edwards, 1998) had data at 33 weeks. Accurate prediction of 3-year outcome depends on longitudinal measurement of sleep behaviors, flexible statistical techniques such as the general linear mixed model, and controlling for the quality of the home environment. Unsupportive home environments can lead to developmental problems even children with good CNS functioning, as shown by the child in Cluster 4, who had a low IQ and hyperactivity.

Because of the exploratory nature of this study, additional research is needed to confirm these findings. In this study, many statistical tests were conducted to determine which potential predictors were most promising. Thus, there is a strong possibility of chance findings. However, the consistent patterning of predictors over developmental domains provides evidence that most findings probably reflect actual relationships.

After replication and refinement, our techniques may make it possible to determine with greater precision the effects of neurological insults during the preterm period. Currently, although the probability of handicap from a specific insult is known, for all but the most severe injuries, some infants show normal development, and some exhibit handicaps. The same insult may have quite different effects on two infants because of differing genetic backgrounds, previous insults, medical and perinatal histories, maturity, time of day, and previous social experiences. Developmental trajectories of sleeping behaviors provide information on how the infant’s brain is functioning despite whatever insults he or she may have experienced.

Finally, many researchers believe that developmental outcome cannot be predicted from the preterm period. Two reasons account for this belief. First, most studies to date have found that social risks are better predictors of outcome than isolated biological risks (Gross et al., 2001; Hack et al., 1992). These studies used rather weak indices of biological risk, usually medical diagnoses. Yet preterm indices, including sucking patterns, EEG energies, neurological risks, and heart rate variability, have been found to relate to long-term outcomes (Brazy, Goldstein, Oehler, Gustafson, & Thompson, 1993; Doussard-Roosevelt et al., 1996; Medoff-Cooper & Gennaro, 1996; Scher et al., 1994). Two studies using these types of biological risks, as well as social risks, found that the effects of social and biological risks were interactive, and the relative effects of risk factors differed at different ages (Smith, Ulvund, & Lindemann, 1994; Thompson et al., 1994). Second, other researchers believe the immature brain is relatively protected in preterms and that except for major insults development is unaffected by preterm events (Parmelee, 1975). However, the development of sleep behaviors during the preterm period, as well as sucking patterns, EEG energies, neurological risks, and heart rate variability, can predict later outcomes. Someday, changes in the developmental trajectories of sleep behaviors may be used, alone or in combination with some of these other measures, to better target early intervention and identify infants with experiences that contributed to either improvements or worsening of their developmental status.

Acknowledgments

We wish to thank Mark Scher for help with conceptualization and Lydia Aydlett, Cathee Huber, Debra H. Brandon, Randall Rieger, Brian Kilgallen, Deborah Assad Lee, Diane Carol Hudson-Barr, Debra B. Miller, and Charlene Garrett for technical assistance. The preparation of this paper was supported by Grant NR01894 from the National Institute for Nursing Research, National Institutes of Health.

References


