Long-term Cognition, Achievement, Socioemotional, and Behavioral Development of Healthy Late-Preterm Infants

Matthew J. Gurka, PhD; Jennifer LoCasale-Crouch, PhD; James A. Blackman, MD, MPH

Objective: To compare healthy late-preterm infants with their full-term counterparts from age 4 through 15 years for numerous standard cognitive, achievement, socioemotional, and behavioral outcomes.

Design: Prospective cohort study.


Participants: A total of 1298 children (53 born at 34-36 weeks’ gestational age), and their families, observed from birth through age 15 years. None of the infants had major health problems before or immediately following birth, and all the infants were discharged from the hospital within 7 days.

Main Exposure: Preterm status: children born late preterm (34-36 weeks) vs those born full term (37-41 weeks).

Main Outcome Measures: Eleven standard outcomes measuring cognition, achievement, social skills, and behavioral/emotional problems using the Woodcock-Johnson Psycho-Educational Battery–Revised and the Child Behavior Checklist, administered repeatedly through age 15 years.

Results: No consistent significant differences were found between late-preterm and full-term children for these standard measures from ages 4 to 15 years. Through age 15 years, the mean difference of most of these outcomes hovered around 0, indicating, along with small confidence intervals around these differences, that it is unlikely that healthy late-preterm infants are at any meaningful disadvantage regarding these measures.

Conclusion: Late-preterm infants born otherwise healthy seem to have no real burdens regarding cognition, achievement, behavior, and socioemotional development throughout childhood.


IN THE PAST 2 DECADES, MUCH ATTENTION has focused on extremely low-birth-weight infants, so-called micropreemies, as the lower limits of survivability have been approached. This work has shown that the lower the gestational age (GA), the higher the risk of death and disability.\(^1\) In addition, research indicates that very preterm infants tend to underachieve academically, have more attention and internalizing behavior problems, and exhibit poor executive function.\(^2\) More recent work\(^3\) examining mortality and morbidity risks in late-preterm infants, defined as those born between 34 and 36 weeks’ GA, has also yielded some concerning results.

The rate of late-preterm infants has risen to 8.8% (in 2003), up from 7.6% a decade earlier.\(^4\) In a hospital-based study\(^5\), spanning 18 months, markedly higher neonatal mortality rates were found for those born at 34 to 36 weeks’ GA compared with those born full term. Another recent study\(^6\) found that infants born at 34 weeks’ GA were 25 times more likely to die in the first week of life and more than 8 times more likely to die in the first month than were infants born at 40 weeks’ GA.

Late-preterm birth has also been associated with significant acute morbidity. For example, compared with term infants, late-preterm infants have more respiratory problems that require mechanical ventilation, jaundice, sepsis, apnea, feeding difficulties, hypoglycemia, temperature instability, and intraventricular hemorrhage.\(^5,7\) In addition, Jain and Cheng\(^8\) found higher rates of emergency department visits for jaundice, feeding problems, infectious issues, and apnea/apparent life-threatening events in late-preterm infants.

Several studies\(^9,11\) with varying methods and sources of data have suggested an increased risk of later developmental and be-
behavioral problems in late-preterm infants, especially as these children enter school. Many of these studies, however, were very large, resulting in significance ($P < 0.05$) of even mild differences, with significant odds ratios (ORs) or risk ratios as low as 1.1. In addition, although being late preterm was found to be a risk factor for cerebral palsy and developmental delay, few studies have focused solely on otherwise healthy late-preterm infants across long periods. Follow-up for graduates of neonatal intensive care has become increasingly short term and focused on the lower end of the GA and birth-weight spectrum, primarily owing to limited resources. Before changes occur in the selection of appropriate candidates for more intensive scrutiny beyond the developmental surveillance recommended in general pediatric practice, further clarification of the extent and nature of the risk is warranted. Using a large longitudinal study, we compared healthy late-preterm infants with their full-term counterparts from ages 4 through 15 years for a variety of cognitive, (academic) achievement, socioemotional, and behavioral outcomes.

### METHODS

#### STUDY DESIGN AND PARTICIPANTS

The National Institute of Child Health and Development Study of Early Child Care and Youth Development (SECCYD) observed children from 10 US sites from birth through age 15 years. Families were recruited at hospital visits after the birth of the child in 1991. Of the 8986 mothers who gave birth during the sampling period, 3416 agreed to participate and were eligible (the mother was aged ≥ 18 years and spoke English, the mother was healthy, the baby was a singleton not given up for adoption, the family lived within 1 hour of the research site, and the neighborhood was sufficiently safe for researchers to visit). Of the eligible mothers, 3015 (56%) were selected at random for a telephone call after 2 weeks using various conditions to ensure representation of single mothers, mothers with less than a high school diploma, and nonwhite mothers. Families were further excluded if the infant had been in the hospital for longer than 7 days, if the baby or the mother was seriously ill, if the family expected to move within 3 years, or if the family could not be reached after 3 contact attempts. In addition to a hospital stay longer than 7 days, baby medical exclusions included being born to a mother known to be addicted to drugs or alcohol, having a chromosomal or genetic abnormality evident at birth that causes severe developmental handicap or disfigurement (eg, Down syndrome and trisomy 18), possessing a congenital defect that causes severe developmental handicap or disfigurement (eg, spina bifida, significant orthopedic handicap, cleft palate, congenital heart disease, deafness, and blindness), having cerebral palsy, having a congenital infection (eg, human immunodeficiency virus, syphilis, rubella, herpes, toxoplasmosis, and cytomegalovirus), and having a genetic or metabolic condition that causes significant developmental handicap not evident in the perinatal period (eg, hypothyroidism and phenylketonuria). Of the remaining 1526 families still eligible, 1364 completed a home interview when the infant was 1 month old. By phase 4 of the study (2005-2007; age 14 and 15 years), 1056 families of the original 1364 (77%) were still enrolled. The GA was calculated based on birth date and due date, as reported by the mother in the hospital; 1348 of the 1364 mothers reported due dates that allowed for calculation of GA.

### CHILD OUTCOMES

#### Cognition/Achievement

Children’s cognition and achievement were assessed using the Woodcock-Johnson Psycho-Educational Battery-Revised, a standardized measure (mean [SD] = 100 [15]) of academic achievement with excellent psychometric properties. At each assessment point, several subtests of the cognitive and achievement batteries were administered, with higher scores indicating improved cognition/achievement. The subtests chosen for this study were administered repeatedly and included Picture Vocabulary, Passage Comprehension, Letter-Word Identification, and Applied Problems.

#### Behavioral/Emotional

The Child Behavior Checklist is a widely used, parent-completed checklist that identifies various types of behavior and emotional problems that occur in children 4 years and older. The measure has been age standardized on large samples of children in the United States and abroad. Each of 118 problem items is scored on a Likert scale based on the preceding 6 months. Scores on each
item are summed to give a raw total problem score, which is then converted to a T-score (mean [SD] = 50 [10]). Higher scores indicate more behavioral and emotional problems. Four of the scales were used in the study to examine behavioral (externalizing, internalizing, and aggressive) and emotional (anxiety/depression) functioning.

Social

Students’ social skills were assessed using the norm-referenced Social Skills Rating System–Teacher Form.19 The social skills composite asks teachers to rate the frequency of classroom behaviors in 3 areas related to positive social adjustment in school settings: cooperation, assertion, and self-control, and a composite score is calculated, with higher scores indicating an improved ability to demonstrate socially accepted learned behaviors.

Relational

Children’s relational functioning was assessed using the Student-Teacher Relationship Scale,20 a 30-item rating scale using a Likert-type format designed to assess teachers’ perceptions of their relationship with a particular student. Two scales, Conflict and Closeness, are calculated and are used as outcomes herein. These scales have been used extensively in studies of school-aged children,21–22 with higher scores indicating more conflict and closeness.

STATISTICAL METHODS

Secondary analysis of the SECCYD data was classified as exempt by the University of Virginia Institutional Review Board for the Social and Behavioral Sciences. All analyses were performed using a statistical software program (SAS version 9.1; SAS Institute Inc, Cary, North Carolina); statistical significance was defined as a P < .05. Late-preterm infants (born at 34-36 weeks’ GA) were compared with full-term infants (born at 37-41 weeks’ GA) on a variety of baseline child and family characteristics. Post-term infants (born at >41 weeks’ GA) were excluded from this comparison. The t test was used for continuous characteristics, such as maternal age, and the χ² test was used to compare binary variable rates, such as sex. Cochran-Armitage tests were used to test for trends when necessary (eg, breastfeeding duration). School referral rates for special services, such as special education, were compared between the 2 groups via the Fisher exact test.

Linear mixed models were used to fit the various continuous outcomes throughout the study. Because these measures were taken during a time when the child was in a certain grade, common times (in months) were assigned for these grade-specific measures (Table 1). The outcomes were modeled as quadratic functions across time (in months), including term status (late preterm vs full term) and its interaction with both time components (linear and quadratic). Thus, this model allowed for the examination of either a constant difference across time between the 2 groups (ie, a difference in the intercept) or a difference in the trajectory across time. Time was centered on 96 months in the model to avoid collinearity. Of primary interest was the estimated difference between the 2 groups across time and its 95% confidence interval (CI). The 95% CI of the model-estimated difference between the 2 groups at any point that did not include 0 indicates statistical significance between late-preterm and full-term infants at that time point.

When statistical significance was not achieved, the 95% CI of the difference estimate was used to determine whether statements of equivalence between the groups could be made.23 That is, the 95% CI provides a range of plausible values of the true difference between late-preterm infants and those born full term. If this range contains differences that one would consider clinically significant, then conclusions regarding equivalence cannot be made. On the other hand, if the range does not contain differences that one would not consider clinically significant, this can be seen as evidence that supports equivalence rather than insufficient statistical power.
Other covariates included in the model as fixed effects were child race (white vs nonwhite), maternal age (in years), maternal education (in years), whether the mother experienced health problems during the pregnancy, delivery type (vaginal or cesarean), mean Home Observation for Measurement of the Environment scores during the first 3 years of life (a measure of the quality of the home environment), mean maternal depression scores (Center for Epidemiological Studies–Depression Scales) during the first 3 years of the child’s life, and the mother’s verbal ability, assessed using the Peabody Picture Vocabulary Test–Revised. To account for the correlation among repeated observations on the same individual, random intercepts and slopes were included in each model. The Kenward-Roger approximation of the degrees of freedom was used to compute $P$ values from the overall tests. These repeated-measures models used all available data on participants, including those who dropped out of the study or had intermittently missing data.

### RESULTS

Of the 1348 children whose GA was collected, 53 (4%) were born before 37 weeks. One was born at 33 weeks’ GA and was excluded from further analysis, as were 47 children born at 42 weeks’ GA and 2 born at 43 weeks’ GA. Late-preterm infants were similar to full-term babies in almost all examined child and family characteristics and in retention in the study (Table 2). The exceptions were maternal age, where late-preterm infants were born to mothers more than 1 year older than their full-term counterparts ($P = .06$), and health problems of the mother during pregnancy, where mothers of late-preterm infants were more likely to have complications such as high blood pres-

---

**Table 3. Gestational Age and Birth Weight Information: Frequency**

<table>
<thead>
<tr>
<th>Participants, No. (%)</th>
<th>Full Term (37-41 wk)</th>
<th>Late Preterm (34-36 wk)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gestational age, wk</strong></td>
<td>(n=1245)</td>
<td>(n=53)</td>
</tr>
<tr>
<td>34</td>
<td>NA</td>
<td>7 (13)</td>
</tr>
<tr>
<td>35</td>
<td>NA</td>
<td>12 (23)</td>
</tr>
<tr>
<td>36</td>
<td>NA</td>
<td>34 (64)</td>
</tr>
<tr>
<td>37</td>
<td>82 (7)</td>
<td>NA</td>
</tr>
<tr>
<td>38</td>
<td>214 (17)</td>
<td>NA</td>
</tr>
<tr>
<td>39</td>
<td>389 (31)</td>
<td>NA</td>
</tr>
<tr>
<td>40</td>
<td>357 (29)</td>
<td>NA</td>
</tr>
<tr>
<td>41</td>
<td>203 (16)</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Birth weight, g</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000-2499</td>
<td>23 (2)</td>
<td>10 (19)</td>
</tr>
<tr>
<td>2500-3999</td>
<td>1033 (83)</td>
<td>43 (81)</td>
</tr>
<tr>
<td>$\geq$4000</td>
<td>189 (15)</td>
<td>0</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.

*One infant born at 33 weeks, 47 born at 42 weeks, and 2 born at 43 weeks were excluded from the study.*
sure and gestational diabetes mellitus \((P = .04)\). Eighty-five percent of the late-preterm infants were born at 35 and 36 weeks’ GA; 7 were born at 34 weeks’ GA \((Table 3)\).

Eighty percent of the late-preterm infants weighed at least 2500 g.

Longitudinal examination of the cognitive and achievement scores through age 15 years revealed no differences between groups, with the mean difference hovering near 0 for all 4 outcomes throughout the entire study period \((Figure 1)\). The 95% CIs around these differences, which provide the range of plausible differences between the groups, show no real possibility of clinically significant lower cognitive/achievement outcomes for those otherwise-healthy late-preterm infants. For the 3 achievement measures, there is a 95% chance that these CIs contain the true difference between the groups, and they generally do not fall below a 5-point disadvantage for the late-preterm children. In fact, the largest absolute limit is approximately 6 points higher.

Late-preterm children have no worse behavioral or emotional problems relative to full-term children \((Figure 2)\). For these 4 standardized measures, the 95% CIs around these estimated differences exclude values greater than a 2-point increase (representing more problematic behavior) for late-preterm infants. Again, the largest limits for these CIs are approximately –4, indicating possibly fewer behavioral problems for preterm infants, if a difference were to exist. In fact, preterm infants were found to have a statistically (but not clinically) significantly lower aggressive score before 5 years, as evidenced by the 95% CI not containing 0.

No differences between late-preterm and full-term children were observed for the social measure of the amount of conflict between the child and the teacher \((Figure 3)\). Although no statistically significant differences were observed for the child’s social skills as rated by his or her teacher(s), the CIs are too wide to prevent the conclusion that these 2 groups are, in fact, equivalent for social skills. The only significant differences were observed in the teacher-child closeness scale, where late-preterm children were found to have higher closeness scores at age 54 months and at 11 years, although the magnitude of these differences was not large (3 and 2 points, respectively).

No differences were observed between school referrals for social services at any time between the first and eighth grades \((Table 4)\). No differences were found for special education referrals, either as a resource or full-time placement. The only difference was in speech/language service referral rates. However, it was the full-term children who were referred more often (17% vs 5% of preterm children, \(P = .03)\). Unlike the primary outcomes measured repeat-

---

**Figure 2.** Estimated differences (late-preterm minus full-term children) in behavioral and emotional measures across time. A, Child Behavior Checklist (CBCL) Externalizing. B, CBCL Internalizing. C, CBCL Aggressive. D, CBCL Anxiety/Depression. A value of 0 at any given time point indicates no difference between groups; a positive difference indicates a higher score for the late-preterm group. Adjusted for sex, race (white vs nonwhite), maternal education (in years), maternal age, delivery type (vaginal vs cesarean), whether the mother had health problems during pregnancy, mean Home Observation for Measurement of the Environment total score (during the first 3 years of life), mean maternal depression score (during the first 3 years of the child’s life), and maternal Peabody Picture Vocabulary Test–Revised score. Solid line indicates mean; broken lines, 95% confidence interval.
The results of this study suggest that late-preterm infants born otherwise healthy have no real disadvantage for cognition, achievement, socioemotional, or behavioral development throughout childhood compared with those born full term. In any study suggesting equivalence, one must address whether the study was sufficiently powered. This was a secondary analysis of an existing data set, so preliminary power analysis for this particular study was not performed. Post hoc power analysis, in general, is not recommended. Rather, the range of possible differences between the groups provided by the 95% CI allows one to conclude whether the suggestion of equivalence is appropriate. If the sample size is too small, this is reflected in a wide CI that includes values that are interpreted as clinically meaningful differences between the groups. For example, if the 95% CI around the mean difference in externalizing behavior score had included 10 (equivalent to 1 SD), then it would be impossible to state that the 2 groups were equivalent for this score because it is possible that the real difference is equal to 10, a clinically meaningful difference.

In the present study, the 95% CIs for the estimated differences generally did not contain clinically meaningful values, at least in the direction in which healthy late-preterm infants would be considered to be at a disadvantage. An average shortfall of 5 points on any Woodcock-Johnson cognitive/achievement score (standardized to have an SD = 15) or of 2 points on any Child Behavior Checklist behavioral measure (standardized to have an SD = 10) is not meaningfully large enough to suggest that healthy late-preterm children have more academic or behavioral problems. In fact, late-preterm children had significantly less aggressive behavior than did full-term children before age 5 years (P < .05) (Figure 2). The observed differences, or the lack of ability to conclude equivalence for the teacher-child closeness measure and the social skills outcome, are worth noting but not highlighting. Given the number of outcomes examined herein, one cannot rule out finding a significant result just by chance alone.

It has been observed that late-preterm infants have more medical problems and increased hospital costs and are at higher risk of cerebral palsy. Other studies have reported statistically significant disadvantages later in childhood for otherwise-healthy late-preterm children with respect to development and achievement. It is important to highlight the distinctions between these studies and the results presented herein. Large sample sizes (>100,000 children) in 2 of the studies allowed for statistical significance even for modest effect sizes (hazard ratio for developmental delay = 1.25, risk ratio for suspension in kindergarten = 1.19). Both were retrospective and, although Morse et al. limited their study to only infants whose hospital stay was 3 days or shorter, the authors acknowledge that some infants may have required readmission after discharge owing to complications such as jaundice and poor feeding. A smaller, but still sizable, prospective study also found statistically significant disadvantages. However, the significant effect sizes were again modest (adjusted ORs generally <1.5). Although a sizable increased odds of special education at participation at kindergarten was observed (adjusted OR = 2.1), the authors defined poor outcomes on reading and math measures as having a score below the 50th percentile. It is difficult to argue that an OR of 1.28 for such an outcome
on a standardized reading score in first grade, although statistically significant, is clinically significant. Gray et al. found clinically significantly higher rates of behavior problems in low-birth-weight infants (<2500 g), but this sample included many infants younger than 34 weeks' GA, whereas 80% of the SECCYD sample of late-preterm infants weighed more than 2500 g.

Although use of the SECCYD in this light presents some clear advantages, including its prospective nature, its long follow-up, and the repeated administration of numerous standard measures to quantify developmental and behavioral constructs, this study is not without weakness. Compared with the other studies characterized herein, this study is relatively small, having only 53 participants classified as late-preterm children. However, as noted previously, this sample size allowed for sufficiently narrow CIs to permit one to conclude that this group is not at a real disadvantage for cognitive, socioemotional, and behavioral development throughout childhood. Nevertheless, larger future studies with similar follow-up would help solidify these conclusions. Furthermore, this study tended to focus on early-childhood (≤3 years old) predictors of these outcomes through adolescence, and a multitude of factors later in a child’s life affect development. Given that most factors were similar between those born late preterm and those born full term (except for maternal age, health during pregnancy, and delivery type), it is doubtful that differences with respect to later-childhood factors could arise between the 2 groups that would confound the (lack of) associations with these outcomes, but it cannot be ruled out definitively.

In addition, the families in this study tended to have higher incomes and education status and agreed to participate in a very long-term study. Thus, this sample may not be representative of the general US population. It is unlikely that this selection bias differs in magnitude between the 2 exposure groups (healthy late preterm vs full term), particularly given that no differences were found between these 2 groups for a host of socioeconomic and other variables (Table 1). However, the exclusionary criteria of this study resulted in a group of late-preterm infants who were relatively healthy, all singletons, with most born between 35 and 36 weeks’ GA (85%) and weighing more than 2500 g (80%). Thus, the results of this study should be generalized only to this healthy subset of late-preterm children.

The results of this study indicate that no real differences exist between healthy late-preterm children and those born full-term for cognition, achievement, socioemotional, and behavioral development from kindergarten through age 15 years. Note that this does not mean that late-preterm infants are not at higher risk for medical problems that may alter these developmental trajectories later in life, and neither do these results endorse early elective delivery of babies. Herein, the key adjective is healthy, as most of these infants weighed more than 2500 g, and all were discharged from the hospital in a timely manner with few or no complications. On the surface, this group may seem too restricted to draw such conclusions. However, being born between 35 and 36 weeks’ GA with no major medical complications is not a rare occurrence. Although being careful in generalizing this rate to the US population, in the SECCYD alone, 53 of 1348 of these otherwise healthy children (4%) were born prematurely. This study, thus, provides comforting information to parents and physicians alike in showing that any real disadvantages of healthy late-preterm children as they mature are minor, if they exist at all. Furthermore, these results suggest that early interventions targeting late-preterm infants may need to focus more on those who have other complications.

Accepted for Publication: December 8, 2009.
Correspondence: Matthew J. Gurka, PhD, Division of Biostatistics and Epidemiology, Department of Public Health Sciences, University of Virginia School of Medicine, PO Box 800717, Charlottesville, VA 22908-0717 (mgurka@virginia.edu).

Author Contributions: Dr Gurka had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Gurka, LoCasale-Crouch, and Blackman. Acquisition of data: Gurka. Analysis and interpretation of data: Gurka, LoCasale-Crouch, and Blackman. Drafting of the manuscript: Gurka, LoCasale-Crouch, and Blackman. Critical revision of the manuscript for important intellectual content: Gurka, LoCasale-Crouch, and Blackman. Statistical analysis: Gurka. Obtained funding: Gurka and Blackman. Administrative, technical, and material support: Gurka. Study supervision: Gurka.

Financial Disclosure: None reported.

Funding/Support: This study was funded by grant R03-HD055298 from the Eunice Kennedy Shriver National Institute of Child Health and Development (Drs Gurka and Blackman).
REFERENCES


Announcement

Trial Registration Required. In concert with the International Committee of Medical Journal Editors (ICMJE), Archives of Pediatrics and Adolescent Medicine will require, as a condition of consideration for publication, registration of all trials in a public trials registry (such as http://ClinicalTrials.gov). Trials must be registered at or before the onset of patient enrollment. This policy applies to any clinical trial starting enrollment after July 1, 2005. The trial registration number should be supplied at the time of submission.

For details about this new policy, and for information on how the ICMJE defines a clinical trial, see the editorials by DeAngelis et al in the September 8, 2004 (2004;292:1363-1364) and June 15, 2005 (2005;293:2927-2929) issues of JAMA. Also see the Instructions to Authors on our Web site: www.archpediatrics.com.