Published in final edited form as:

Paediatr Perinat Epidemiol. 2016 November ; 30(6): 571–582. doi:10.1111/ppe.12319.

Describing the shape of the relationship between gestational age at birth and cognitive development in a nationally representative U.S. birth cohort

Jennifer L. Richards^a, Carolyn Drews-Botsch^a, Jessica M. Sales^b, W. Dana Flanders^a, and Michael R. Kramer^a

^a Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

^b Department of Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Abstract

Background—Preterm children face higher risk of cognitive and academic deficits compared with their full-term peers. The objective of this study was to describe early childhood cognitive ability and kindergarten academic achievement across gestational age at birth in a population-based longitudinal cohort.

Methods—The study population included singletons born at 24–42 weeks GA enrolled in the Early Childhood Longitudinal Study-Birth Cohort (n=6,150 for 2-year outcome, n=4,450 for kindergarten outcome). Home-based assessments measured cognitive ability at 2 years and reading and mathematics achievement at kindergarten age. Linear regression models estimated the association between gestational age and cognitive and academic scores using four different ways of modeling gestational age: continuous variable in linear and quadratic terms; categories for individual weeks; and clinical categories for early preterm, moderate preterm, late preterm, early term, full term, late term, and post-term.

Results—Children born at early preterm (24–27 weeks), moderate preterm (28–33 weeks), and late preterm (34–36 weeks) scored significantly worse than full-term (39–40 weeks) peers on 2-year and kindergarten assessments; however, no deficits were observed for early term (37–38 weeks). These categories were a clinically useful and parsimonious approach to stratifying risk of adverse cognitive and academic outcomes.

Conclusions—This study estimated the relative performance of children born at 24–42 weeks in a population-based birth cohort using multiple approaches to modeling gestational age, providing a more rigorous understanding of the relationships between the full spectrum of gestational age and cognitive and academic outcomes in early childhood and at school age.

Keywords

Gestational age; Preterm birth; Cognitive and academic outcomes; Birth cohort

INTRODUCTION

Children who are born preterm are more likely to experience cognitive and academic deficits as well as learning disabilities and problems in school.^{1,2} Preterm delivery interrupts *in utero* brain growth and development; children are born with developmentally immature brains before they have had sufficient time to achieve optimal size and neuronal development. Preterm children may have smaller brain volumes^{3,4} and experience central nervous system injuries^{5,6}, both of which are associated with worsened neurodevelopmental outcomes in childhood. Children born at the earliest gestational ages and lowest birth weights have the most severe consequences; however, even children born nearer to term—i.e., at late preterm or early term—may have worsened cognitive and academic outcomes compared those born at full term or at 39 weeks or later.^{7–10}

Numerous studies have evaluated children's outcomes following preterm or early term birth but few have examined outcomes along the entire range of gestational age within the same analytic population. Given evidence of the adverse impacts of even mild prematurity, we wanted to understand the impacts of each additional week spent *in utero* on cognitive scores or school outcomes without limiting our analyses by using a single reference group (e.g., 37 weeks). Variation in cognitive test scores by each additional gestational week has been found among children born earlier than 33 weeks¹¹ and among infants born at 37–41 weeks.^{12–14} Two studies examined the relationship between the full range of gestational age and academic problems, focusing on special education provisions in Scotland¹⁵ and completion of basic schooling requirements in Denmark.¹⁶ The objective of this study was to describe cognitive and academic outcomes at 2 years and kindergarten age across the full spectrum of gestational age at birth among children enrolled in a U.S. population-based longitudinal cohort study.

METHODS

This study used data from the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), a population-based longitudinal study that sampled children from 2001 U.S. births and followed them through kindergarten. Its design and data collection procedures have been described extensively.¹⁷ Briefly, more than 14,000 children were selected in a stratified, clustered sample of 2001 birth certificates with oversampling of twins, low birth weight, very low birth weight, and certain racial/ethnic groups. About 10,700 children including 8,850 singletons enrolled in the study at 9 months old. Data were also collected at 2 years, preschool age, and kindergarten age. Kindergarten data collection was split across two waves starting in fall 2006 and fall 2007 in order to assess children at kindergarten entry, which typically differed based on birth date; we used data from children's first entry into kindergarten. The kindergarten waves followed an 85% sub-sample (n=7,700; 6,250 singletons) of children who had completed the preschool wave (n=9,000; 7,400 singletons). To be eligible for follow-up in each subsequent wave, children needed to have a completed parent interview at the previous wave.

There were 7,250 singletons born at 24-42 completed weeks who were eligible for this study. Children were excluded if part of a multiple birth ($n=1,800$), missing clinical estimate of gestational age ($n=1,450$) or birth weight ($n=150$), or reported birth weight was implausible for gestational age ($n<50$).¹⁸ Nearly all (97%) of children missing clinical estimate of gestational age were California births, where the clinical estimate was not reported on birth certificates until adoption of the 2003 revised birth certificate.¹⁹

Gestational age at delivery was defined using the clinical estimate of gestation reported on the birth certificate. Gestational age was modeled using four methods spanning three general approaches: continuous (linear and quadratic terms); using each gestational age week as its own category; and using categories defined by the American Congress of Obstetricians and Gynecologists (ACOG) for early preterm (24-27 weeks), moderate preterm (28-33 weeks), late preterm (34-36 weeks), early term (37-38 weeks), term (39-40 weeks), late term (41 weeks), and post-term (42 weeks).²⁰

Cognitive ability at 2 years old was measured using the Bayley Short Form-Research Edition (BSF-R) Mental Scale, an abbreviated version of the Bayley Scales of Infant Development-Second Edition (BSID-II) developed for the ECLS-B.²¹ The BSID-II comprises 178 items organized into age sets; individual test takers receive different sets of items based on their assessment age (corrected for prematurity). In contrast, ECLS-B children were all administered the same set of up to 33 items—including 19 core items, 5 basal items, and 9 ceiling items—selected based on psychometric properties to validly test cognitive ability for 22-26 month old children. Our analyses used children's scale scores, which represented the item response theory (IRT)-based estimate of raw score on the full BSID-II.²² Some children completed the BSF-R Mental Scale outside of the target age range of 22-26 months; therefore we restricted to children in this age range in a sensitivity analysis.

Reading and mathematics achievement at kindergarten age was measured using ECLS-B-designed assessments of knowledge and skills in reading and mathematics.¹⁷ These assessments used existing items from standardized instruments and assessment batteries for preschool- and kindergarten-aged children (e.g., Peabody Picture Vocabulary Test, PreLAS® 2000), as well as ECLS-B-developed items.²³ There was a two-stage design consisting of a routing test administered to all children and a second-stage test of low, medium, or high difficulty selected based on routing test performance. Our analyses used children's scale scores, which represented the IRT-based estimate of the number of items that a child would have answered correctly if they had been asked all of the scored questions.

Covariates of interest were child's race/ethnicity (American Indian or Alaska Native, Asian, non-Hispanic black, Hispanic, multi-racial, Native Hawaiian or other Pacific Islander, non-Hispanic white), maternal educational attainment (less than high school, completed high school, some college, bachelor's degree or higher), household poverty adjusted for household size ($<130\%$ of the federal poverty threshold [FPL], 130% to $<185\%$ of FPL, and 185% of FPL), child's sex, parity (first-born, second or third born, fourth-born or above), and maternal age at delivery (categorized as 15-17, 18-19, 20-24, 25-29, 30-34, 35-39, and 40 years or older). Covariates were ascertained from birth certificates (child's sex, parity, maternal age at delivery) or parent interviews at 9 months (child's race/ethnicity, maternal

educational attainment, household poverty for 2-year analyses) and 2 years (household poverty for kindergarten analyses).

Statistical analysis

Descriptive statistics included frequencies by gestational age and covariates as well as unadjusted mean outcome scores by categories of gestational age and covariates. Multivariable analyses with generalized estimating equations (GEE) based on the normal distribution estimated mean differences in outcome scores at each time point, accounting for clustering of subjects by ECLS-B designed sample clusters. Sampling weights were not applied in the primary analyses because the objective was to obtain internally valid estimates of the relationship between gestational age and outcome scores among study participants rather than to ensure that findings were representative of 2001 U.S. births.

In order to describe the shape of the relationship between gestational age and each outcome, four methods of modeling gestational age were compared: (1) categories for each week comparing to a 40-week referent category, (2) ACOG categories comparing to a full term (39-40 weeks) referent category, (3) continuous (linear model), and (4) continuous adding a quadratic term (i.e., squared gestational age). Model fit was compared using quasi-likelihood under the independence model (QIC) statistics analogous to Akaike information criterion (AIC) statistics for generalized estimating equations. Cubic terms were also considered but dropped from final analyses due to limited statistical power especially in the smaller gestational age categories and because they did not offer superior model fit compared to quadratic models. To graphically compare across gestational age definitions, predicted scores for each week of gestation were plotted at covariate reference values.

All analyses were adjusted for child's assessment age. Clinical recommendations suggest correcting for prematurity until two or three years of age; therefore we used corrected age for the primary 2-year analyses. Correcting for prematurity means that a child's degree of prematurity (e.g., 3 weeks premature) is subtracted from his or her chronological age measured from birth date in order to arrive at a "testing age" for developmental assessments—effectively resulting in a premature child being compared with term-born children of higher chronological age. Typical practice is to subtract the number of weeks born early (compared with 37 weeks) from the child's chronological age.²⁴ Given recent evidence of the impacts of mild prematurity on cognitive outcomes, however, we corrected children's testing ages to a referent of 40 weeks, and we applied the correction for all children born earlier than 40 weeks. For example, a child born at 36 weeks and tested at 24 months chronological age was assigned a correcting age of 23 months. Sensitivity analyses were conducted using chronological age for the BSF-R Mental Scale outcome, and using prematurity-corrected age for the kindergarten outcomes. Analyses were also adjusted for child's race/ethnicity, maternal educational attainment, household poverty, child's sex, parity, and maternal age at delivery.

About 8% (n = 550) of eligible children who completed the 2-year wave were missing BSF-R Mental Scale scores. In a sensitivity analysis, missing BSF-R scores were multiply imputed using multivariate imputation by chained equations.²⁵ One-third of the eligible sample was lost to follow up between the 2-year and kindergarten waves. In a sensitivity

analysis, the possibility of selection bias influencing the kindergarten results was evaluated using inverse probability of censorship weighting (IPCW) and marginal structural models.^{26,27} Among eligible children who completed the 2-year wave (n=6,700), logistic regression was used to obtain predicted probabilities of participating in the kindergarten wave based on baseline covariates. In weighted GEE models for the kindergarten outcomes, non-censored children were assigned stabilized inverse probability weights, effectively up-weighting subjects who were less likely to complete the kindergarten wave.

Analyses were conducted using SAS v9.3 (Cary, NC) and R 3.1 (R Foundation for Statistical Computing, Vienna, Austria). This study was approved by the Emory Institutional Review Board. Unweighted frequencies are rounded to the nearest 50 per National Center for Education Statistics guidelines.

RESULTS

A total of 6,150 children were included in 2-year analyses and 4,450 were included in kindergarten analyses (eFigure 1). Due to oversampling of low birth weight children, about 18% were born before 37 weeks (Table 1). Children included in our analyses were racially and socioeconomically diverse: 57% were racial/ethnic minorities, 50% lived under 185% of the FPL, and 50% had mothers who had completed a high school degree or less. The mean BSF-R Mental Scale score was 126.1 (SD=11.0, range 92.4-174.1). Mean scores on the kindergarten reading and mathematics assessments were 44.4 (SD=14.8, range 12.4-82.5) and 43.9 (SD=10.5, range 11.2-69.7) respectively.

Children born at 24-39 weeks performed worse compared with children born at 40 weeks, although estimated deficits were not statistically significant for those born at 31 or 35-39 weeks after covariate adjustment (Table 2; Figure 1A). Using ACOG categories, significant deficits were observed for the early preterm (-6.6, 95% CI: -8.1, -5.0), moderate preterm (-2.9, 95% CI: -4.0, -1.8), and late preterm (-1.3, 95% CI: -2.3, -0.4) groups. Using a linear term for continuous gestational age suggested a significant, positive relationship between gestational age and BSF-R Mental Scale score. The quadratic function of continuous gestational age was significant but very small in magnitude. There was little difference in QIC statistics across adjusted models. When models were instead adjusted for chronological age at assessment, estimated deficits were stronger (eTable 1). Significant deficits were observed for the early preterm (-12.9, 95% CI: -14.2, -11.5), moderate preterm (-7.2, 95% CI: -8.2, -6.3), late preterm (-3.3, 95% CI: -4.2, -2.5), as well as early term (-1.1, 95% CI: -1.7, -0.5) groups.

Children born at late preterm or earlier scored significantly worse on the kindergarten reading and mathematics assessments compared to those born at term (Tables 3-4, Figures 1B-1C). Using individual week categories, most groups under 37 weeks exhibited significant deficits in reading and mathematics scale scores compared to children born at 40 weeks. Using ACOG categories, significant reading deficits were observed for early preterm (-5.1, 95% CI: -7.2, -3.0), moderate preterm (-3.0, 95% CI: -4.3, -1.8), and late preterm (-1.8, 95% CI: -3.3, -0.4). On the mathematics assessment, significant deficits were observed for early preterm (-6.7, 95% CI: -8.5, -4.9), moderate preterm (-3.6, 95% CI: -4.6, -2.7), and

late preterm (−1.6, 95% CI: −2.6, −0.6). For both outcomes, we found that only the linear term was significant in models operationalizing gestational age as a continuous variable; quadratic terms were not significant. There was little difference in QIC comparing across adjusted models for each of the kindergarten outcomes. When adjusted for corrected age, estimated deficits for preterm and early term children on both the reading and mathematics assessments were attenuated (eTables 2-3). For reading, estimated deficits were no longer significant for early preterm, moderate preterm, or late preterm compared with term births. For mathematics, significant deficits remained for early preterm (−4.4, 95% CI: −6.2, −2.6) and moderate preterm (−2.0, 95% CI: −3.0, −1.1).

Factors associated with missing 2-year scores were early preterm and moderate preterm delivery, household poverty, hospital stays due to medical problems after birth, receipt of early intervention or other support services for special needs at 9 months or 2 years, and having a parent-reported health problem or impairment that limited ability to walk, run, or play. Rates of chromosomal anomalies or birth defects were similar across children with and without 2-year scores. Results were similar when missing 2-year scores (n=550) were multiply imputed (eTable 4). When 2-year analyses were restricted to children tested at 22-26 months of age, results were similar in models adjusting for chronological age (eTable 5). In models adjusting for corrected age, estimated deficits for early preterm and moderate preterm children were slightly attenuated. Larger proportions of these groups were excluded based on the 22-26 month criterion after correcting children's ages for prematurity.

About half of those lost to follow up between the 2-year and kindergarten waves were part of the planned 85% sample size reduction between the preschool and kindergarten waves (eFigure 1). These children were more likely to be non-Hispanic white, be preterm, live in poverty, and have younger and less educated mothers, but did not differ in terms of 2-year scores. Applying IPCW stabilized weights to account for potential selection bias did not meaningfully change results for the kindergarten outcomes (eTables 6-7).

COMMENT

Children born at early preterm, moderate preterm, and late preterm had deficits in cognitive ability at 2 years old and reading and mathematics achievement scores at kindergarten age compared with their peers born at full term. Magnitude of deficits increased with decreasing gestational age.

Our examination of the relationship between gestational age at delivery and cognitive and academic outcomes in early childhood using multiple approaches to modeling the full continuum of gestational age contributes to the extant literature that often dichotomizes or coarsely categorizes preterm births in some way (e.g., <32 weeks, 34-36 weeks) and uses 37 weeks or later as a reference group. The consequence of these categorical schemes is that children with different risk may be inappropriately pooled into common groups—our objective was to formally test the assumption that such categories adequately describe outcome risks. These analyses of multiple methods for modeling gestational age support the use of ACOG categories—which were developed primarily to capture differences in neonatal medical outcomes—as a useful, parsimonious approach to capturing overall

patterns in cognitive and academic outcomes by gestational age. Similar fit statistics across models suggested that none of the methods for operationalizing gestational age resulted in a model that clearly had comparatively superior predictive power. For all three outcomes, the use of ACOG categories adequately described the shape of the relationship between gestational age and cognitive/academic score within our study cohort and did so with clinically meaningful cut-points.

This study benefitted from the use of data from a large birth cohort with rigorous prospective data collection and direct assessment of children's cognitive and academic outcomes. The population-based design of the study, however, comes with some inherent limitations, namely that even with over-sampling of low birth weight infants and a large overall sample, numbers of children at earlier gestational weeks, especially below 32 weeks, were still relatively small. There may have been low power to detect associations and identify non-linear relationships when categorizing gestational age by individual week and using continuous functions of gestational age, particularly in very preterm range where non-linearities might be observed. Being able to potentially delineate differences in outcomes within the early ACOG categories (e.g., very preterm spanning 24-27 weeks and moderate preterm spanning 28-33 weeks), with comparison to outcomes along the full spectrum of gestational age as we have done in this study, may be useful for neonatologists and pediatricians as well as for parents and families of very and moderately preterm children. Additional studies with larger sample sizes at extreme gestational weeks is needed to estimate week-by-week differences, and we suggest that these studies use a similar approach to ours analyzing the full spectrum of gestational age to better understand the distribution of outcomes rather than restricting to pre-specified reference groups.

Further, a limitation of using population-based secondary data—in contrast to, for example, follow up of hospital-based populations—is potential underrepresentation of children with severe disabilities that preclude them and their families from agreeing to participate. In this study, we did find that children with parent-reported impairments or who received early intervention services were slightly less likely to complete the 2-year assessment. Our results should be interpreted with the caveat that they may be representative of the group of preterm and early term children who are well-functioning enough to participate in a longitudinal study and to complete standard cognitive and academic assessments.

The extant literature suggests that higher risk of cognitive and academic problems and educational problems persists up through early term births,^{28,29} but we found a significant deficit for ECLS-B children born at early term only on the BSF-R Mental Scale when adjusting for chronological instead of corrected age. It is possible that this was driven by correcting for prematurity for all children born earlier than 40 weeks rather than the more common clinical practice of applying the correction if born earlier than 37 weeks. Further examination of the appropriate gestational age cut-off for applying a “prematurity” correction is needed in light of more recent findings of cognitive and academic deficits nearer to term.

It is also interesting to note that adjusting for corrected age for the kindergarten outcomes substantially attenuated estimated deficits for preterm children. This may suggest—in line

with several recent studies^{30–32}—that correction for prematurity may be warranted beyond two or three years of age, an issue that remains under debate. Use of a prematurity correction relies on an assumption that developmental outcomes of preterm children temporarily lag behind full-term children due to their early delivery and therefore shorter time since conception for central nervous system maturation. Clinical recommendations to correct for prematurity through two or three years old are based on the notion that preterm children are expected to eventually catch up with their term peers as the pace of developmental growth slows. The use of chronological age is supported by the rationale that environmental factors such as the home environment, medical care, and parent-child interactions also play an important role in post-birth development.³³ In practical terms, chronological age may be used in screening for early intervention services and becomes even more relevant as children enter school based on their birth date. A recent survey of pediatric health care providers in a primary care network in Pennsylvania, New Jersey, and Delaware, showed that chronological age was used in developmental surveillance in 71% of visits for children born <32 weeks³⁴ suggesting the importance of understanding potential differences in observed outcomes when using corrected versus chronological age.

A few additional limitations should be mentioned. Our analyses used the clinical estimate of gestational age due to measurement concerns related to using LMP dating³⁵ and therefore excluded California births; results were similar, however, when analyses were repeated using gestational age based on LMP dating. Measurement error in gestational age reported on the birth certificate is possible and may potentially vary with factors such as socioeconomic status and timing of entry into prenatal care. Further, the analytic sample dropped by about 30% between Wave 2 and the kindergarten waves raising concerns about potential selection bias. There was little change in our findings when IPCW weights were applied to account for differential probabilities of loss to follow up. Longitudinal trajectories of cognitive and academic scores were not studied because the BSF-R Mental Scale used at age 2 years and the ECLS-B kindergarten reading and mathematics were not designed to be comparable assessments.

A large body of literature demonstrates that preterm and early term children are at higher risk of cognitive deficits and worsened academic outcomes in childhood. This study contributes to that literature by estimating the relative performance of children at all gestational ages in a population-based birth cohort using multiple approaches to modeling gestational age, providing a more rigorous understanding of the relationships between the full spectrum of gestational age and cognitive and academic outcomes in early childhood and at school age.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Jennifer Richards received support from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH) T32 Predoctoral Training Program in Reproductive, Perinatal, and Pediatric Epidemiology under Award Number T32HD052460. Michael R. Kramer

received support from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH) under Award Number K01-HD074726. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Ms. Richards had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

- Behrman, RE.; Butler, AS., editors. Preterm Birth: Causes, Consequences, and Prevention. National Academies Press (US); Washington (DC): 2007. Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. <http://www.ncbi.nlm.nih.gov/books/NBK11362/>. Accessed August 18, 2014
- Allen MC. Neurodevelopmental outcomes of preterm infants. *Curr Opin Neurol*. 2008; 21(2):123–128. [PubMed: 18317268]
- Keunen K, Kersbergen KJ, Groenendaal F, Isgum I, de Vries LS, Benders MJNL. Brain tissue volumes in preterm infants: prematurity, perinatal risk factors and neurodevelopmental outcome: A systematic review. *J Matern Fetal Neonatal Med*. 2012; 25(S1):89–100.
- Peterson BS, Anderson AW, Ehrenkranz R, et al. Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. *Pediatrics*. 2003; 111(5):939–948. Pt 1. [PubMed: 12728069]
- Neubauer A-P, Voss W, Kattner E. Outcome of extremely low birth weight survivors at school age: the influence of perinatal parameters on neurodevelopment. *Eur J Pediatr*. 2008; 167(1):87–95. [PubMed: 17333273]
- Luu TM, Ment LR, Schneider KC, Katz KH, Allan WC, Vohr BR. Lasting effects of preterm birth and neonatal brain hemorrhage at 12 years of age. *Pediatrics*. 2009; 123(3):1037–1044. [PubMed: 19255037]
- Chyi LJ, Lee HC, Hintz SR, Gould JB, Sutcliffe TL. School outcomes of late preterm infants: special needs and challenges for infants born at 32 to 36 weeks gestation. *J Pediatr*. 2008; 153(1):25–31. [PubMed: 18571530]
- Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. *Pediatrics*. 2009; 123(4):e622–629. [PubMed: 19336353]
- Stephens BE, Vohr BR. Neurodevelopmental outcome of the premature infant. *Pediatr Clin North Am*. 2009; 56(3):631–646. [PubMed: 19501696]
- Poulsen G, Wolke D, Kurinczuk JJ, et al. Gestational age and cognitive ability in early childhood: a population-based cohort study. *Paediatr Perinat Epidemiol*. 2013; 27(4):371–379. [PubMed: 23772939]
- Johnson S. Cognitive and behavioural outcomes following very preterm birth. *Semin Fetal Neonatal Med*. 2007; 12(5):363–373. [PubMed: 17625996]
- Yang S, Platt RW, Kramer MS. Variation in child cognitive ability by week of gestation among healthy term births. *Am J Epidemiol*. 2010; 171(4):399–406. [PubMed: 20080810]
- Noble KG, Fifer WP, Rauh VA, Nomura Y, Andrews HF. Academic achievement varies with gestational age among children born at term. *Pediatrics*. 2012; 130(2):e257–e264. [PubMed: 22753563]
- Rose O, Blanco E, Martinez SM, et al. Developmental scores at 1 year with increasing gestational age, 37–41 weeks. *Pediatrics*. 2013; 131(5):e1475–e1481. [PubMed: 23589812]
- MacKay DF, Smith GCS, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS Med*. 2010; 7(6):e1000289. [PubMed: 20543995]
- Mathiasen R, Hansen BM, Andersen A-MNN, Forman JL, Greisen G. Gestational age and basic school achievements: a national follow-up study in Denmark. *Pediatrics*. 2010; 126(6):e1553–1561. [PubMed: 21059721]
- National Center for Education Statistics. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Kindergarten 2006 and 2007: Data File User's Manual (2010-010). 2010
- Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national reference for fetal growth. *Obstet Gynecol*. 1996; 87(2):163–168. [PubMed: 8559516]

19. Martin JA, Osterman MJ, Kirmeyer SE, Gregory EC. Measuring gestational age in vital statistics data: transitioning to the obstetric estimate. National Vital Statistics Reports. 64 Number 5. http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_05.pdf.
20. ACOG Committee Opinion No. 579: Definition of term pregnancy. Obstet Gynecol. 2013; 122(5): 1139–1140. [PubMed: 24150030]
21. National Center for Education Statistics. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), User's Manual for the ECLS-B Longitudinal 9-month-2-year Data File and Electronic Codebook. May.2006
22. Andreassen, C.; Fletcher, P.; Park, J. Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) Psychometric Report for the 2-Year Data Collection. National Center for Education Statistics; 2007. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007084>
23. National Center for Education Statistics. The ECLS-B Direct Assessment: Choosing the Appropriate Score for Analysis. <http://nces.ed.gov/ecls/pdf/birth/ChoosingScores.pdf>. Accessed March 11, 2015
24. Committee on Fetus and Newborn. Age Terminology During the Perinatal Period. Pediatrics. 2004; 114(5):1362–1364. [PubMed: 15520122]
25. vaan Buren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011; 45(3) <https://stat.ethz.ch/education/semesters/ss2012/ams/paper/mice.pdf>. Accessed July 18, 2016.
26. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology. 2016; 27(1):91–97. [PubMed: 26484424]
27. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168(6):656–664. [PubMed: 18682488]
28. Dong Y, Chen S, Yu J. A systematic review and meta-analysis of long-term development of early term infants. Neonatology. 2012; 102(3):212–221. [PubMed: 22814228]
29. Quigley MA, Poulsen G, Boyle E, et al. Early term and late preterm birth are associated with poorer school performance at age 5 years: a cohort study. Arch Dis Child Fetal Neonatal Ed. 2012; 97(3):F167–173. [PubMed: 22215800]
30. Wilson-Ching M, Pascoe L, Doyle LW, Anderson PJ. Effects of correcting for prematurity on cognitive test scores in childhood. J Paediatr Child Health. 2014; 50(3):182–188. [PubMed: 24617343]
31. Parekh SA, Boyle EM, Guy A, et al. Correcting for prematurity affects developmental test scores in infants born late and moderately preterm. Early Hum Dev. 2016; 94:1–6. [PubMed: 26826320]
32. van Veen S, Aarnoudse-Moens CSH, van Kaam AH, Oosterlaan J, van Wassenae-Leemhuis AG. Consequences of correcting intelligence quotient for prematurity at age 5 years. J Pediatr. 2016; 173:90–95. [PubMed: 26979649]
33. Wilson SL, Craddock MM. Review: Accounting for prematurity in developmental assessment and the use of age-adjusted scores. J Pediatr Psychol. 2004; 29(8):641–649. [PubMed: 15491987]
34. D'Agostino JA, Gerdes M, Hoffman C, Manning ML, Phalen A, Bernbaum J. Provider use of corrected age during health supervision visits for premature infants. J Pediatr Health Care. 2013; 27(3):172–179. [PubMed: 23611457]
35. Ananth CV. Menstrual versus clinical estimate of gestational age dating in the United States: temporal trends and variability in indices of perinatal outcomes. Paediatr Perinat Epidemiol. 2007; 21:22–30. [PubMed: 17803615]

Figure 1A-C. Model-predicted BayleyShort Form-Research Edition (BSF-R) Mental Scale scores at age 2 years (1A), kindergarten reading scale scores (1B), and kindergarten mathematics scale scores (1C), by gestational age

Predicted scores are for a hypothetical individual at covariate reference values (child race/ethnicity = non-Hispanic white, maternal education = bachelor's degree or higher, household poverty = 185 percent of federal poverty level, child's sex = female, parity = nulliparous, maternal age at birth = 25-29 years) taking each assessment at the median age of testing (corrected 24 months for 2-year outcome, 68 months for kindergarten outcomes). Shaded region of figure denotes full term referent group (39-40 weeks).

Table 1

Characteristics of the analytic cohorts.

	2-year cognitive scores				Kindergarten scores					
	BSF-R Mental Scale				Reading			Mathematics		
	N	(%)	Mean	SD	N	(%)	Mean	SD	Mean	SD
Total	6,150		126.1	11.0	4,450		44.4	14.8	43.9	10.5
Gestational age (weeks)										
24	50	(0.4)	113.5	10.1	*	(0.4)	38.2	13.4	35.2	11.3
25	50	(0.5)	110.9	11.6	*	(0.5)	40.9	16.8	37.6	11.5
26	50	(0.8)	115.7	8.4	50	(0.8)	37.7	13.6	35.4	10.8
27	100	(1.4)	114.4	9.5	50	(1.3)	42.4	15.3	41.1	9.8
28	100	(1.5)	117.5	10.2	50	(1.5)	40.9	13.7	39.9	9.2
29	100	(1.2)	118.5	10.2	50	(1.4)	38.0	13.5	38.1	9.2
30	100	(1.4)	119.0	10.6	50	(1.3)	41.7	16.2	39.9	11.7
31	50	(1.1)	122.9	8.6	50	(1.1)	44.1	13.1	41.6	10.1
32	50	(0.9)	120.3	11.2	50	(0.8)	42.3	13.2	40.6	10.9
33	100	(1.2)	121.2	9.5	50	(1.3)	44.4	14.4	43.7	9.6
34	100	(1.8)	121.0	9.9	50	(1.5)	42.9	14.2	42.4	10.5
35	150	(2.1)	125.4	11.9	100	(1.9)	42.5	14.9	41.7	10.1
36	250	(4.1)	124.5	9.7	150	(3.8)	42.0	14.4	42.3	10.6
37	450	(7.6)	125.5	11.7	350	(7.7)	44.1	14.7	43.9	10.5
38	1,000	(16.2)	126.6	10.4	750	(16.6)	44.3	15.0	44.1	10.7
39	1,450	(23.7)	127.4	10.6	1,050	(23.9)	45.1	14.9	44.8	10.3
40	1,550	(25.2)	128.2	10.5	1,150	(25.7)	45.9	14.8	44.9	10.2
41	500	(7.7)	128.0	10.3	350	(7.7)	44.0	14.6	44.5	10.7
42	50	(0.9)	127.3	9.5	50	(0.8)	45.5	15.2	45.4	8.8)
Child's race/ethnicity										
Black or African American, non-Hispanic	1,100	(17.8)	123.0	10.5	800	(18.5)	41.3	14.2	40.2	9.9)
Hispanic	1,050	(17.4)	122.6	10.5	800	(17.6)	39.9	14.2	40.6	9.7)
Asian	550	(9.1)	126.4	11.6	450	(10.6)	54.0	16.3	50.3	10.3)
Native Hawaiian or other Pacific Islander	50	(0.5)	121.9	10.3	*	(0.3)	35.9	9.8	36.5	6.9

	2-year cognitive scores					Kindergarten scores				
	BSF-R Mental Scale					Reading				
	N	(%)	Mean	SD		N	(%)	Mean	SD	Mathematics
American Indian or Alaska Native	200	(2.9)	124.9	9.1		150	(3.6)	38.7	12.9	39.9
More than 1 race, non-Hispanic	550	(8.8)	127.2	10.7		400	(9.2)	45.3	14.8	45.3
White, non-Hispanic	2,650	(43.3)	128.6	10.7		1,750	(40.0)	45.7	13.8	45.6
Missing	*	(0.2)				*	(0.3)			9.9
Assessment age (months), mean (SD)	24.5	(1.3)				68.3	(4.2)			
Corrected assessment age (months), mean (SD)	23.9	(1.6)				67.8	(4.3)			
Child's sex										
Female	3,000	(48.8)	127.9	10.6		2,200	(49.4)	45.5	14.2	44.2
Male	3,150	(51.2)	124.3	11.0		2,250	(50.6)	43.3	15.3	43.7
Maternal age at birth (years)										
15-17	250	(4.3)	123.7	10.1		200	(3.9)	37.4	12.6	38.3
18-19	500	(8.5)	123.9	10.1		350	(8.1)	38.6	13.8	39.8
20-24	1,650	(26.7)	125.0	10.3		1,150	(26.1)	41.9	14.2	41.8
25-29	1,500	(24.3)	126.6	11.1		1,100	(24.4)	45.4	14.6	44.6
30-34	1,400	(22.6)	127.5	11.6		1,050	(23.7)	47.6	15.2	46.5
35-39	700	(11.1)	127.0	11.1		500	(11.2)	47.7	14.6	46.5
40	150	(2.6)	127.1	11.9		100	(2.6)	46.8	13.4	46.7
Parity										
Nulliparous	2,550	(41.6)	126.5	11.2		1,850	(42.3)	45.7	15.1	44.3
1-2 previous live births	2,950	(48.3)	126.4	10.8		2,150	(48.0)	44.5	14.5	44.3
3+ previous live births	600	(9.7)	123.1	10.5		400	(9.4)	38.4	14.5	40.3
Missing	50	(0.4)				*				10.8
Maternal education										
Less than high school	1,150	(18.6)	122.6	9.9		800	(17.7)	36.4	13.5	38.0
Completed high school	1,900	(31.2)	124.5	10.7		1,350	(30.5)	41.1	13.6	41.4
Some college	1,500	(24.6)	126.5	10.9		1,100	(24.6)	44.9	13.4	44.6
Bachelor's degree or higher	1,550	(25.5)	130.2	10.9		1,200	(27.1)	53.0	14.0	50.1
Missing	*	(0.1)				*				8.9
Household poverty at 9 months										

	2-year cognitive scores				Kindergarten scores			
	BSF-R Mental Scale				Reading		Mathematics	
	N	(%)	Mean	SD	N	(%)	Mean	SD
<130% of federal poverty level	2,250	(36.8)	123.2	10.3				
130 to <185% of federal poverty level	800	(12.7)	124.8	10.8				
185% of federal poverty level	3,100	(50.4)	128.5	10.9				
Household poverty at 2 years								
<130% of federal poverty level					1,500	(33.9)	38.6	14.0
130 to <185% of federal poverty level					550	(12.4)	41.7	13.3
185% of federal poverty level					2,400	(53.7)	48.7	14.3
							47.1	9.7

NOTE: Unweighted sample sizes are rounded to the nearest 50 per National Center for Education Statistics guidelines. Frequencies denoted * round to zero. SD = standard deviation.

Table 2

Association between gestational age and Bayley Short Form-Research Edition (BSF-R) Mental Scale score at 2 years old

Gestational age	Adjusted for corrected age	Fully adjusted*
	β (95% CI)	β (95% CI)
Categorical (1-week categories)		
24	-7.9 (-11.9, -4.0)	-5.8 (-9.6, -2.1)
25	-10.5 (-14.0, -7.0)	-8.8 (-11.9, -5.7)
26	-6.9 (-9.0, -4.8)	-5.4 (-7.5, -3.2)
27	-8.4 (-10.3, -6.4)	-7.1 (-8.9, -5.3)
28	-5.2 (-7.2, -3.2)	-3.8 (-5.7, -1.8)
29	-5.2 (-7.6, -2.8)	-4.0 (-6.1, -1.8)
30	-4.9 (-7.3, -2.6)	-3.8 (-5.8, -1.8)
31	-1.7 (-3.5, 0.1)	-0.6 (-2.5, 1.2)
32	-4.5 (-7.6, -1.4)	-3.4 (-6.6, -0.2)
33	-3.7 (-5.9, -1.5)	-2.8 (-4.9, -0.6)
34	-4.5 (-6.5, -2.6)	-3.6 (-5.4, -1.9)
35	-0.6 (-2.7, 1.4)	-0.3 (-2.3, 1.6)
36	-2.0 (-3.4, -0.6)	-1.1 (-2.5, 0.2)
37	-1.3 (-2.4, -0.2)	-0.8 (-1.9, 0.2)
38	-0.5 (-1.3, 0.3)	-0.1 (-0.8, 0.7)
39	-0.4 (-1.0, 0.2)	-0.3 (-0.9, 0.4)
40	0.0 (Reference)	0.0 (Reference)
41	-0.1 (-1.3, 1.0)	0.0 (-1.0, 1.0)
42	-1.2 (-3.7, 1.2)	-0.2 (-2.4, 2.0)
ACOG categories		
Early preterm (22-27 weeks)	-8.0 (-9.5, -6.4)	-6.6 (-8.1, -5.0)
Moderate preterm (28-33 weeks)	-4.0 (-5.1, -2.9)	-2.9 (-4.0, -1.8)
Late preterm (34-36 weeks)	-2.0 (-3.0, -0.9)	-1.3 (-2.3, -0.4)
Early term (37-38 weeks)	-0.5 (-1.2, 0.1)	-0.2 (-0.8, 0.4)
Term (39-40 weeks)	0.0 (Reference)	0.0 (Reference)
Late term (41 weeks)	0.0 (-1.0, 1.1)	0.1 (-0.9, 1.1)
Post term (42 weeks)	-1.1 (-3.5, 1.4)	-0.1 (-2.2, 2.0)
Continuous: Linear model		
1 week increase in GA	0.5 (0.4, 0.6)	0.4 (0.3, 0.5)
Continuous: Quadratic model		
1 week increase in GA	1.9 (0.7, 3.1)	1.8 (0.7, 3.0)
Squared term	-0.02 (-0.04, 0.00)	-0.02 (-0.04, 0.00)

n = 6,150. Mean (SD) score on the BSF-R Mental Scale was 126.1 (11.0). β estimates change in BSF-R Mental Scale score.

* Adjusted for child's corrected age at assessment, child's race/ethnicity, child's sex, maternal age at birth, parity, maternal educational attainment, and household poverty.

Table 3

Association between gestational age and kindergarten reading scale score.

Gestational age	Adjusted for chronological age	Fully adjusted*
	β (95% CI)	β (95% CI)
Categorical (1-week categories)		
24	-9.8 (-15.6, -4.0)	-7.5 (-13.7, -1.3)
25	-7.2 (-11.8, -2.5)	-7.8 (-12.4, -3.2)
26	-9.5 (-13.4, -5.7)	-7.0 (-10.0, -4.0)
27	-4.6 (-8.1, -1.1)	-3.2 (-6.6, 0.1)
28	-6.4 (-9.2, -3.5)	-4.4 (-6.9, -1.8)
29	-8.9 (-12.2, -5.6)	-7.3 (-10.0, -4.5)
30	-4.7 (-8.0, -1.4)	-2.6 (-5.7, 0.5)
31	-2.3 (-6.2, 1.5)	-1.7 (-5.2, 1.9)
32	-4.9 (-9.0, -0.7)	-3.4 (-6.8, 0.1)
33	-2.2 (-5.7, 1.3)	-0.8 (-3.9, 2.4)
34	-3.1 (-6.1, -0.1)	-1.5 (-4.3, 1.3)
35	-4.0 (-6.8, -1.2)	-2.8 (-5.5, -0.2)
36	-4.2 (-6.5, -2.0)	-2.3 (-4.1, -0.5)
37	-1.5 (-3.1, 0.1)	-1.4 (-2.8, 0.0)
38	-1.2 (-2.6, 0.2)	-0.8 (-2.0, 0.3)
39	-1.1 (-2.3, 0.0)	-0.9 (-2.0, 0.1)
40	0.0 (Reference)	0.0 (Reference)
41	-0.6 (-2.5, 1.2)	-0.2 (-1.7, 1.3)
42	0.0 (-4.7, 4.8)	0.9 (-3.3, 5.2)
ACOG categories		
Early preterm (22-27 weeks)	-6.5 (-8.8, -4.3)	-5.1 (-7.2, -3.0)
Moderate preterm (28-33 weeks)	-4.5 (-5.9, -3.1)	-3.0 (-4.3, -1.8)
Late preterm (34-36 weeks)	-3.4 (-5.0, -1.8)	-1.8 (-3.2, -0.4)
Early term (37-38 weeks)	-0.8 (-1.9, 0.3)	-0.5 (-1.5, 0.4)
Term (39-40 weeks)	0.0 (Reference)	0.0 (Reference)
Late term (41 weeks)	-0.1 (-1.9, 1.7)	0.2 (-1.2, 1.7)
Post term (42 weeks)	0.6 (-4.2, 5.4)	1.4 (-2.9, 5.6)
Continuous: Linear model		
1 week increase in GA	0.5 (0.4, 0.6)	0.4 (0.3, 0.5)
Continuous: Quadratic model		
1 week increase in GA	0.6 (-1.3, 2.6)	0.8 (-0.8, 2.4)
Squared term	0.00 (-0.03, 0.03)	-0.01 (-0.03, 0.02)

n = 4,450. Mean (SD) scale score on kindergarten reading assessment was 44.4 (14.8). β estimates change in kindergarten reading scale score.

* Adjusted for child's chronological age at assessment, child's race/ethnicity, child's sex, maternal age at birth, parity, maternal educational attainment, and household poverty.

Table 4

Association between gestational age and kindergarten mathematics scale score.

Gestational age	Adjusted for chronological age	Fully adjusted*
	β (95% CI)	β (95% CI)
Categorical (1-week categories)		
24	-10.8 (-16.0, -5.7)	-9.4 (-14.9, -3.8)
25	-9.4 (-13.2, -5.7)	-9.5 (-13.8, -5.3)
26	-10.1 (-13.3, -6.9)	-8.6 (-11.4, -5.8)
27	-4.8 (-7.2, -2.5)	-4.0 (-6.4, -1.6)
28	-5.9 (-8.0, -3.9)	-4.3 (-6.3, -2.3)
29	-7.7 (-9.8, -5.6)	-6.5 (-8.6, -4.5)
30	-5.2 (-7.8, -2.6)	-3.8 (-6.2, -1.4)
31	-3.5 (-6.3, -0.6)	-2.6 (-5.2, 0.0)
32	-5.1 (-8.4, -1.7)	-4.1 (-6.9, -1.2)
33	-1.9 (-4.3, 0.5)	-0.9 (-2.9, 1.2)
34	-2.7 (-4.9, -0.4)	-1.4 (-3.6, 0.8)
35	-3.4 (-5.4, -1.3)	-2.6 (-4.5, -0.7)
36	-2.7 (-4.2, -1.2)	-1.4 (-2.7, -0.1)
37	-0.7 (-1.8, 0.4)	-0.6 (-1.6, 0.4)
38	-0.5 (-1.3, 0.4)	-0.2 (-1.0, 0.5)
39	-0.3 (-1.0, 0.5)	-0.2 (-0.9, 0.4)
40	0.0 (Reference)	0.0 (Reference)
41	0.5 (-0.8, 1.9)	0.9 (-0.2, 2.0)
42	1.1 (-1.7, 3.9)	1.6 (-0.9, 4.2)
ACOG categories		
Early preterm (22-27 weeks)	-7.7 (-9.5, -5.9)	-6.7 (-8.5, -4.9)
Moderate preterm (28-33 weeks)	-4.8 (-5.8, -3.8)	-3.6 (-4.6, -2.7)
Late preterm (34-36 weeks)	-2.8 (-3.9, -1.6)	-1.6 (-2.6, -0.6)
Early term (37-38 weeks)	-0.4 (-1.1, 0.3)	-0.2 (-0.9, 0.4)
Term (39-40 weeks)	0.0 (Reference)	0.0 (Reference)
Late term (41 weeks)	0.6 (-0.6, 1.9)	1.0 (0.0, 2.1)
Post term (42 weeks)	1.3 (-1.5, 4.0)	1.8 (-0.8, 4.3)
Continuous: Linear model		
1 week increase in GA	0.6 (0.5, 0.6)	0.5 (0.4, 0.5)
Continuous: Quadratic model		
1 week increase in GA	1.7 (0.3, 3.1)	1.7 (0.5, 3.0)
Squared term	-0.02 (-0.04, 0.00)	-0.02 (-0.04, 0.00)

n = 4,450. Mean (SD) scale score on kindergarten mathematics assessment was 43.9 (10.5). β estimates change in kindergarten mathematics scale score.

* Adjusted for child's chronological age at assessment, child's race/ethnicity, child's sex, maternal age at birth, parity, maternal educational attainment, and household poverty.