



Neurocognitive Development

Gestational age at term delivery and children's neurocognitive development

Jessica L Gleason ¹, Stephen E Gilman ^{2,3}, Rajeshwari Sundaram,⁴ Edwina Yeung,¹ Diane L Putnick,¹ Yassaman Vafai,¹ Abhisek Saha⁴ and Katherine L Grantz ^{1*}

¹Epidemiology Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA, ²Social and Behavioral Sciences Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA, ³Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and ⁴Biostatistics Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

*Corresponding author. Epidemiology Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, 6710B Rockledge Dr., Bethesda, MD 20892, USA.
E-mail: katherine.grantz@nih.gov

Received 27 January 2021; editorial decision 30 May 2021; Accepted 16 June 2021

Abstract

Background: Preterm birth is associated with lower neurocognitive performance. However, whether children's neurodevelopment improves with longer gestations within the full-term range (37–41 weeks) is unclear. Given the high rate of obstetric intervention in the USA, it is critical to determine whether long-term outcomes differ for children delivered at each week of term.

Methods: This secondary analysis included 39 199 live-born singleton children of women who were admitted to the hospital in spontaneous labour from the US Collaborative Perinatal Project (1959–76). At each week of term gestation, we evaluated development at 8 months using the Bayley Scales of Infant Development, 4 years using the Stanford–Binet IQ (SBIQ) domains and 7 years using the Wechsler Intelligence Scales for Children (WISC) and Wide-Range Achievement Tests (WRAT).

Results: Children's neurocognitive performance improved with each week of gestation from 37 weeks, peaking at 40 or 41 weeks. Relative to those delivered at 40 weeks, children had lower neurocognitive scores at 37 and 38 weeks for all assessments except SBIQ and WISC Performance IQ. Children delivered at 39 weeks had lower Bayley Mental ($\beta = -1.18$; confidence interval $-1.77, -0.58$) and Psychomotor ($\beta = -1.18$; confidence interval $-1.90, -0.46$) scores. Results were similar for within-family analyses comparing siblings, with the addition of lower WRAT scores at 39 weeks.

Conclusions: The improvement in development scores across assessment periods indicates that each week up to 40 or 41 weeks of gestation is important for short- and long-term cognitive development, suggesting 40–41 weeks may be the ideal delivery window for optimal neurodevelopmental outcomes.

Key words: Neurocognitive development, term delivery, child development, obstetrics

Key Messages

- Previous studies show that children's neurocognition improves when delivered in the late vs early term period (39–41 vs 37–38 weeks of gestation).
- Little is known about children's long-term neurocognition when delivered at each week of term gestation.
- Using data from the US Collaborative Perinatal Project, in assessments at 8 months, 4 years and 7 years, children's neurocognitive development peaked when delivered at 40 or 41 weeks.
- Results remained consistent when adjusting for potential confounders, including within-family effects and when accounting for potential errors in gestational dating.

Introduction

Being born preterm is associated with lower neurocognitive performance. This association holds for children who are born extremely or very preterm, but also for children delivered in the late preterm period between 34 and 36 weeks, who carry an elevated risk for cognitive delay.¹ Development of the cerebral cortex continues into the term period, so longer gestation confers an advantage for brain development. Because brain development continues through the weeks that encompass term and late-term delivery (37–41 weeks of gestation),² researchers have begun to evaluate the cognitive development of children delivered between 37 and 41 weeks. However, most studies group term births into a large heterogeneous reference group.^{3–15} Only one study has evaluated differences by individual week of term delivery, finding that longer gestation through to 41 weeks was associated with higher development scores, although that study was limited by its small sample size and only evaluated outcomes through to 12 months of age.¹⁶ No other study has evaluated differences in the cognitive development of children born at each successive week of gestation within the range of term delivery and none has considered cognitive development at multiple follow-up points in childhood.

Given the high rate of obstetric interventions during pregnancy in the USA and a recent practice advisory from the American College of Obstetricians and Gynecologists supporting elective induction at 39 weeks of gestation even in low-risk pregnancies,¹⁷ the question of whether longer gestation is better for children born at term has significant

implications for obstetric practice. Understanding differences in development by week is also critical to assist in clinical decision-making, as the decision on when to deliver is made on a week-by-week basis. Accordingly, we examined differences in the neurocognitive development of children from infancy through to early childhood who were born at 37, 38, 39, 40, 41 and ≥ 42 weeks of gestation.

Methods

The Collaborative Perinatal Project (CPP) was a US prospective cohort study that recruited pregnant women from 12 clinical centres between 1959 and 1966.¹⁸ Of the 46 021 women enrolled, there were 59 391 pregnancies, with $\sim 54\,902$ live births. Children were followed to age 8 years, receiving detailed assessments of neurocognitive development, general health and physical growth. The study has been described in detail elsewhere.¹⁸ For this study, we included all live-born singletons ($n = 53\,647$), excluding children delivered prior to 37 weeks of gestation ($n = 8844$), women who were admitted for delivery not in spontaneous labour ($n = 5134$) and infants who died prior to the 8-month assessment ($n = 470$), leaving a final analytic sample of 39 199 children.

Gestational age

Women were interviewed at enrolment, where they reported their last menstrual period (LMP), which was used to determine gestational age (GA) at delivery. GA was defined by 1-week intervals of 37 (i.e. included 37 weeks

0 days through to 37 weeks 6 days), 38, 39, 40, 41 and 42+ (combined 42–45 weeks).

Neurocognitive outcomes

Children underwent detailed neurocognitive assessments by trained clinicians at ages 8 months, 4 years and 7 years. All examinations followed a standardized protocol with extensive quality-control procedures. Assessments administered were reliable and validated within the age groups tested and are currently in use in updated forms for contemporary developmental testing.¹⁸ To assess infant development, CPP psychologists administered the Bayley Scales of Infant Development Mental and Psychomotor exams at 8 months. Bayley raw scores were age-standardized. As part of the assessment, psychologists could assign a rating of 'normal', 'suspect' or 'abnormal'. We dichotomized results for both Bayley Scales into 'normal' or 'suspect/abnormal' for binary analyses.

At 4 years, clinicians administered the Stanford–Binet IQ (SBIQ) test. At 7 years, psychologists administered the Wechsler Intelligence Scale for Children (WISC) and Wide-Range Achievement Test (WRAT). We used results from the WISC Full Scale IQ, Verbal IQ, Performance IQ, as well as WRAT Reading, Spelling, and Arithmetic as primary outcomes for age 7 years. All developmental assessments were normalized, such that a score of 100 equates to average. Children who scored <85 (one standard deviation below the mean) on any of the 4- or 7-year assessments were classified as having below-average intelligence or achievement.

Covariates

During the enrolment interview, women provided their age (years), parity (0, 1, 2, 3+ children) and marital status (single, married/living with partner or divorced/widowed/separated). They also reported information on their and the father's educational attainment, occupational status and household income that were used to derive the socio-economic index of the CPP (0–1, 2–3.9, 4–5.9, 6–7.9, 8–9.9), which is interpreted as a percentile score of socio-economic status.¹⁹ Smoking was assessed at enrolment and each prenatal visit as to whether they ever smoked in this pregnancy (yes or no) and the number of cigarettes smoked per day (1–5, 6–10, 11–20, >20). For this analysis, we combined these into an ever smoke during pregnancy category (yes or no). Additionally, information to diagnose hypertensive diseases of pregnancy (gestational hypertension and pre-eclampsia) was abstracted from medical charts. Gestational hypertension was defined as at least two elevated blood-pressure readings across prenatal care visits

(systolic >140 or diastolic >90 mmHg) at or after 20 weeks of gestation or a report of gestational hypertension in the delivery summary. Pre-eclampsia was defined as gestational hypertension with proteinuria, headache or visual disturbances within 1 week of delivery or pulmonary oedema.

Child sex and race were reported by clinicians during child neurocognitive assessments. The site at which participants were enrolled was also included as a covariate for analyses.

Analysis

Descriptive statistics for each covariate were calculated to compare participants across each week of GA at delivery using chi-squared tests for categorical variables and Kruskal–Wallis tests for continuous variables. Because some participants did not attend all three assessment visits, some assessments were missing at 8 months ($n = 6291$), 4 years ($n = 10\,100$) and 7 years ($n = 9168$). To account for potential selection bias due to missing assessments, we calculated stabilized inverse-probability weights (IPWs) for each age-grouping of outcomes to account for the conditional probability of having completed the assessment. Estimates using these weights can be interpreted as the association between GA at delivery and neurocognitive outcomes that would have been observed if all children had completed the assessment for the specified age (8 months, 4 years or 7 years). We used all covariates described previously in the calculation of weights and all analyses were weighted by calculated IPW. Delivery at 40 weeks was the reference group for all analyses.

To test the association between GA at delivery and each continuous neurocognitive score, we fit unadjusted and adjusted generalized linear models. Because some women in the CPP had multiple children, we used generalized estimating equations to account for sibling clusters. Using these regression models, we calculated adjusted means for each outcome across weeks of GA at delivery. We also performed a within-family analysis to partially account for hereditary factors and other unmeasured variables that would cluster among families. In analyses that included only siblings, we fit linear mixed models with a random intercept by maternal ID to estimate within-family differences in outcomes, adjusting for factors that would differ between siblings—birth order and sex. Maternal smoking and hypertensive diseases may change between pregnancies but were similar between siblings in this sample.

Additionally, we calculated the relative risk (RR) of scoring suspect or abnormal on the Bayley Scales or below average (<85) on all other assessments by fitting log-linear models assuming a Poisson distribution. For these models,

we also accounted for sibling clusters using generalized estimating equations.

Sensitivity analyses

To test the robustness of our results in consideration of using LMP to estimate GA in the absence of ultrasound dating information, we conducted a sensitivity analysis to account for the error in recall in self-reported LMP. We simulated data with errors in recall from a symmetric distribution (mean = 1, standard deviation = 8)^{20,21} constrained between 36.5 and 43.5 weeks. Details of the simulation are given in the [Supplementary Methods](#), available as [Supplementary data](#) at *IJE* online. In an additional

sensitivity analysis, we repeated the main analyses among a subsample of women ($n = 37\,497$) excluding those with gestational hypertension, pre-eclampsia or gestational diabetes, as these conditions may influence placental sufficiency and subsequent neurocognitive outcomes.

Role of the funding source

The funding source for this work had no influence on the collection, analysis or interpretation of the data, nor did it influence the writing of the report or decision on where to submit the paper for publication.

Table 1. Maternal and child characteristics by gestational age at delivery, Collaborative Perinatal Project ($n = 39\,199$)

	Full sample ^a	Gestational age at delivery (weeks)					
		37	38	39	40	41	42+
Total [n (%)]	39 199	3028 (7.7)	5570 (14.2)	8705 (22.2)	9370 (23.9)	6104 (15.6)	6422 (16.4)
Maternal characteristics							
Age [mean (SD)]	24.1 (5.9)	23.7 (6.2)	24.1 (5.9)	24.2 (5.9)	24.2 (5.8)	24.0 (5.9)	23.8 (5.9)
Socio-economic index [n (%)]							
0–1.9	2802 (7.4)	290 (9.9)	436 (8.1)	619 (7.3)	576 (6.3)	384 (6.5)	497 (8.0)
2–3.9	11 838 (31.1)	1080 (36.9)	1770 (32.7)	2544 (30.1)	2634 (28.8)	1666 (28.1)	2144 (34.6)
4–5.9	11 839 (31.1)	945 (32.2)	1740 (32.2)	2640 (31.2)	2823 (30.9)	1776 (30.0)	1915 (30.9)
6–7.9	7609 (19.9)	425 (14.5)	1007 (18.6)	1744 (20.6)	1978 (21.7)	1346 (22.7)	1109 (17.9)
8–9.9	3979 (10.5)	191 (6.5)	456 (8.4)	915 (10.8)	1122 (12.3)	756 (12.8)	539 (8.7)
Parity [n (%)]							
0	12 278 (31.7)	922 (30.9)	1672 (30.3)	2648 (30.7)	2985 (32.1)	2043 (33.8)	2008 (31.7)
1	8913 (23.0)	661 (22.1)	1280 (23.2)	2042 (23.7)	2191 (23.6)	1385 (22.9)	1354 (21.4)
2	6188 (15.9)	491 (16.4)	900 (16.3)	1423 (16.5)	1430 (15.4)	918 (15.2)	1026 (16.2)
3+	11 415 (29.4)	913 (30.6)	1666 (30.2)	2508 (29.1)	2682 (28.9)	1697 (28.1)	1949 (30.8)
Marital status [n (%)]							
Single	5652 (14.4)	556 (18.4)	894 (16.1)	1202 (13.8)	1236 (13.2)	805 (13.2)	959 (14.9)
Married/living with partner	30 405 (77.6)	2222 (73.4)	4241 (76.1)	6873 (79.0)	7438 (79.4)	4818 (78.9)	4813 (75.0)
Divorced, widowed, separated	3142 (8.0)	250 (8.3)	435 (7.8)	630 (7.2)	696 (7.4)	481 (7.9)	650 (10.1)
Ever smoked this pregnancy [n (%)]	23 526 (60.7)	1752 (58.7)	3293 (59.7)	5130 (59.5)	5696 (61.4)	3720 (61.8)	3935 (62.2)
Gestational hypertension [n (%)]	708 (1.8)	44 (1.5)	94 (1.7)	125 (1.4)	188 (2.0)	134 (2.2)	123 (1.9)
Pre-eclampsia [n (%)]	940 (2.4)	81 (2.7)	162 (2.9)	189 (2.2)	221 (2.4)	129 (2.1)	158 (2.5)
Child characteristics							
Sex, male [n (%)]	19 444 (49.6)	1408 (46.5)	2689 (48.3)	4276 (49.1)	4657 (49.7)	3205 (52.5)	3209 (50.0)
Race/ethnicity [n (%)]							
White	18 628 (47.5)	1010 (33.4)	2104 (37.8)	3222 (42.4)	4931 (52.6)	3433 (56.2)	3205 (49.9)
Black	17 557 (44.8)	1747 (57.7)	3050 (54.8)	3834 (50.4)	3792 (40.5)	2225 (36.5)	2652 (41.3)
Asian	196 (0.5)	13 (0.4)	45 (0.8)	49 (0.6)	39 (0.4)	30 (0.5)	20 (0.3)
Puerto Rican	2508 (6.4)	231 (7.6)	294 (6.7)	559 (6.4)	529 (5.7)	361 (5.9)	490 (7.6)
Other	309 (0.8)	27 (0.9)	33 (0.8)	61 (0.7)	78 (0.8)	55 (0.9)	55 (0.9)

^aLiveborn singletons born at 37+ weeks, admitted to hospital in spontaneous labour, excluding infant death <8 months.

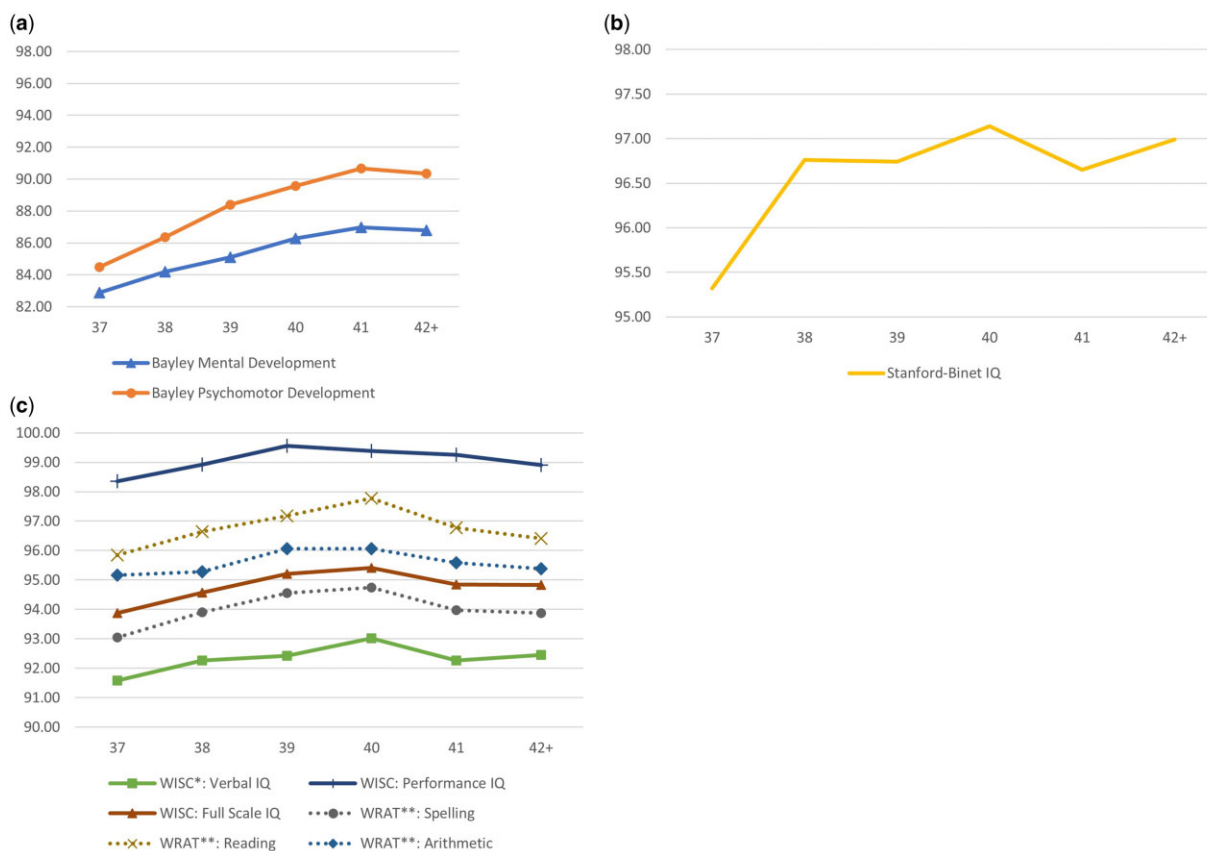


Figure 1. Adjusted mean scores for 8-month ($n = 32\,908$) (A), 4-year ($n = 29\,099$) (B) and 7-year ($n = 30\,031$) (C) assessments by gestational age at delivery. Scores represent the population mean for each week of gestation, calculated from main analytic models (linear models with generalized estimating equations for sibling clusters), adjusted for maternal age, socio-economic index, tobacco use, parity, pre-eclampsia, gestational hypertension, child race and study site. WISC, Wechsler Intelligence Scales for Children; WRAT, Wide-Range Achievement Test.

Results

Demographic and other characteristics of the full sample can be viewed in [Table 1](#). The sample primarily consisted of married (77.6%), multiparous women (68.3%) with a mean age of 24 years (SD = 5.9). The majority of their children were identified as White (47.5%) or Black (44.8%), with a small percentage from Asian, Puerto Rican or other race/ethnic groups. There were differences across GA groups for all covariates except pre-eclampsia.

Children's neurocognitive test scores increased with each week of gestation from 37 weeks through to 40 or 41 weeks and declined in the post-term period (42+ weeks) ([Figure 1](#)). In main adjusted analyses where delivery at 40 weeks was the reference group across all outcomes, at 8 months, there were linear decreasing trends in the scores on both Bayley Scales from 37 to 39 weeks ([Table 2](#)), with improvements in scores for delivery at 41 weeks for Mental [$\beta = 0.70$; 95% confidence interval (CI) 0.02, 1.37] and Psychomotor ($\beta = 1.11$; 95% CI 0.31, 1.91) development. Similar linear patterns were observed for 4- and 7-year

assessments, though scores were highest at 40 weeks. At the 7-year assessment, children delivered at 38 weeks had lower WISC verbal ($\beta = -0.75$; CI $-1.43, -0.06$) and Full Scale IQ ($\beta = -0.83$; CI $-1.54, -0.13$), as well as lower WRAT Spelling ($\beta = -0.83$; CI $-1.45, -0.20$), Reading ($\beta = -1.14$; CI $-1.91, -0.37$) and Arithmetic ($\beta = -0.78$; CI $-1.37, -0.20$) scores. In within-family models restricted to siblings, results were similar to main analyses in terms of magnitude, direction and significance, though there were many additional deficits observed in the post-term period (42+ weeks) ([Table 3](#)). Compared with their siblings born at 40 weeks, in addition to lower scores across all outcomes for 37 and 38 weeks, children born at 39 weeks had lower SBIQ ($\beta = -1.13$; CI $-2.01, -0.25$) scores and lower scores on WRAT Spelling ($\beta = -0.75$; CI $-1.46, -0.05$), Reading ($\beta = -1.12$; CI $-1.96, -0.28$) and Arithmetic ($\beta = -0.68$; CI $-1.31, -0.05$) tests.

In analyses of below-average and suspect/abnormal neurodevelopment, children delivered at 37 and 38 weeks had higher risks of having below-average scores on almost all domains across assessment times ([Table 4](#)). Those

Table 2. Associations between gestational age at delivery and neurocognitive outcomes showing mean difference in scores from reference (40 weeks) (*n* = 39 199)

	37	38	39	40	41	42+
8-month						
Bayley Mental	-3.39 (-4.28, -2.49)	-2.10 (-2.77, -1.42)	-1.18 (-1.77, -0.58)	[Reference]	0.70 (0.02, 1.37)	0.52 (-0.13, 1.17)
Bayley Psychomotor	-5.08 (-6.14, -4.02)	-3.22 (-4.03, -2.40)	-1.18 (-1.90, -0.46)	[Reference]	1.11 (0.31, 1.91)	0.77 (-0.02, 1.56)
4-year						
SBIQ	-1.83 (-2.86, -0.79)	-0.38 (-1.18, 0.41)	-0.41 (-1.12, 0.30)	[Reference]	-0.49 (-1.29, 0.30)	-0.15 (-0.92, 0.62)
7-year						
WISC: VIQ	-1.43 (-2.34, -0.52)	-0.75 (-1.43, -0.06)	-0.60 (-1.22, 0.02)	[Reference]	-0.75 (-1.42, -0.08)	-0.56 (-1.24, 0.11)
WISC: PIQ	-1.03 (-2.02, -0.03)	-0.47 (-1.24, 0.30)	0.17 (-0.50, 0.85)	[Reference]	-0.14 (-0.88, 0.60)	-0.48 (-1.24, 0.27)
WISC: FSIQ	-1.54 (-2.49, -0.59)	-0.83 (-1.54, -0.13)	-0.19 (-0.82, 0.43)	[Reference]	-0.55 (-1.24, 0.14)	-0.58 (-1.27, 0.11)
WRAT: Spelling	-1.69 (-2.51, -0.87)	-0.83 (-1.45, -0.20)	-0.19 (-0.76, 0.38)	[Reference]	-0.77 (-1.39, -0.14)	-0.87 (-1.48, -0.26)
Reading	-1.93 (-2.91, -0.95)	-1.14 (-1.91, -0.37)	-0.60 (-1.31, 0.11)	[Reference]	-1.01 (-1.78, -0.23)	-1.37 (-2.13, -0.62)
Arithmetic	-0.89 (-1.67, -0.12)	-0.78 (-1.37, -0.20)	0.01 (-0.52, 0.53)	[Reference]	-0.48 (-1.05, 0.09)	-0.68 (-1.27, -0.09)

Results of linear models with generalized estimating equations for sibling clusters and robust standard error, adjusted for maternal age, socio-economic index, tobacco use, parity, pre-eclampsia, gestational hypertension, child race and study site, presented as β ; 95% confidence interval. FSIQ, Full Scale IQ; PIQ, performance IQ; SBIQ, Stanford-Binet Intelligence (IQ) test; VIQ, verbal IQ; WISC, Wechsler Intelligence Scales for Children; WRAT, Wide-Range Achievement Test.

Table 3. Associations between gestational age at delivery and neurocognitive outcomes for within-family effect sibling models showing mean difference in developmental scores from reference (40 weeks)

	37	38	39	40	41	42+
8-month						
Bayley Mental	-3.26 (-4.40, -2.11)	-2.80 (-3.70, -1.91)	-1.63 (-2.40, -0.87)	[Reference]	0.65 (-0.18, 1.48)	-0.27 (-1.13, 0.59)
Bayley Psychomotor	-4.33 (-5.68, -2.99)	-3.19 (-4.24, -2.14)	-1.27 (-2.19, -0.34)	[Reference]	0.58 (-0.39, 1.55)	0.77 (-0.25, 1.78)
4-year						
SBIQ	-2.45 (-3.77, -1.14)	-1.93 (-2.97, -0.89)	-1.13 (-2.01, -0.25)	[Reference]	-1.0 (-1.94, -0.06)	-1.66 (-2.65, -0.66)
7-year						
WISC: VIQ	-1.95 (-3.04, -0.87)	-1.03 (-1.87, -0.18)	-0.50 (-1.22, 0.22)	[Reference]	0.41 (-0.36, 1.19)	-1.05 (-1.86, -0.24)
WISC: PIQ	-2.75 (-4.00, -1.50)	-1.13 (-2.10, -0.15)	-0.49 (-1.33, 0.35)	[Reference]	-0.07 (-0.97, 0.83)	-1.13 (-2.06, -0.19)
WISC: FSIQ	-2.51 (-3.63, -1.38)	-1.09 (-1.97, -0.22)	-0.63 (-1.38, 0.12)	[Reference]	0.31 (-0.50, 1.11)	-1.06 (-1.90, -0.22)
WRAT: Spelling	-2.56 (-3.61, -1.50)	-0.99 (-1.81, -0.16)	-0.75 (-1.46, -0.05)	[Reference]	-0.31 (-1.07, 0.45)	-1.67 (-2.46, -0.88)
Reading	-2.53 (-3.78, -1.27)	-1.29 (-2.27, -0.31)	-1.12 (-1.96, -0.28)	[Reference]	0.79 (-1.70, 0.11)	-2.26 (-3.20, -1.31)
Arithmetic	-1.40 (-2.34, -0.46)	-0.80 (-1.54, -0.06)	-0.68 (-1.31, -0.05)	[Reference]	-0.24 (-0.91, 0.44)	-1.24 (-1.94, -0.53)

Restricted to participants with sibling in the Collaborative Perinatal Project (*n* = 10 350). Within-family effect sibling models adjusted for birth order and child sex, presented as β ; 95% confidence interval. FSIQ, Full Scale IQ; PIQ, performance IQ; SBIQ, Stanford-Binet Intelligence (IQ) test; VIQ, verbal IQ; WISC, Wechsler Intelligence Scales for Children; WRAT, Wide-Range Achievement Test.

Table 4. Adjusted relative risk of scoring below average or suspect/abnormal on neurocognitive assessments according to gestational week at delivery

		37	38	39	40	41	42+
8-month	Bayley Mental	1.67 (1.34, 2.08)	1.37 (1.13, 1.66)	1.05 (0.88, 1.27)	[Reference]	0.98 (0.79, 1.21)	0.86 (0.69, 1.07)
	Bayley Psychomotor	1.81 (1.43, 2.29)	1.56 (1.28, 1.91)	1.17 (0.96, 1.42)	[Reference]	0.97 (0.77, 1.21)	0.94 (0.75, 1.18)
4-year	SBIQ	1.19 (1.05, 1.36)	1.02 (0.91, 1.14)	1.01 (0.91, 1.12)	[Reference]	1.06 (0.94, 1.19)	1.01 (0.90, 1.13)
7-year	WISC: VIQ	1.12 (1.01, 1.25)	1.07 (0.97, 1.17)	1.02 (0.93, 1.11)	[Reference]	1.04 (0.94, 1.15)	1.06 (0.96, 1.16)
	WISC: PIQ	1.16 (1.01, 1.33)	1.13 (1.01, 1.27)	0.98 (0.87, 1.10)	[Reference]	1.06 (0.94, 1.20)	1.11 (0.98, 1.25)
	WISC: FSIQ	1.21 (1.08, 1.35)	1.12 (1.01, 1.24)	1.03 (0.94, 1.14)	[Reference]	1.10 (0.99, 1.22)	1.11 (1.00, 1.22)
	WRAT: Spelling	1.31 (1.15, 1.48)	1.14 (1.02, 1.28)	1.00 (0.90, 1.12)	[Reference]	1.11 (0.99, 1.24)	1.05 (0.94, 1.18)
	Reading	1.26 (1.08, 1.48)	1.10 (0.95, 1.27)	1.09 (0.96, 1.24)	[Reference]	1.13 (0.98, 1.30)	1.05 (0.92, 1.21)
	Arithmetic	1.15 (0.99, 1.34)	1.15 (1.01, 1.31)	1.08 (0.97, 1.27)	[Reference]	1.11 (0.97, 1.27)	1.16 (1.02, 1.32)

Results of log-linear models assuming a Poisson distribution and robust standard error with generalized estimating equations for sibling clusters, adjusted for maternal age, socio-economic index, tobacco use, parity, pre-eclampsia, gestational hypertension, child race and study site, presented as RR; 95% confidence interval. FSIQ, Full Scale IQ; PIQ, performance IQ; SBIQ, Stanford-Binet Intelligence (IQ) test; VIQ, verbal IQ; WISC, Wechsler Intelligence Scales for Children; WRAT, Wide-Range Achievement Test.

delivered at 38 weeks had a higher risk of being classified as suspect or abnormal on either Bayley Mental (RR = 1.37; CI 1.13, 1.66) and Psychomotor (RR = 1.56; CI 1.28, 1.91) and higher risk of scoring below average on WISC Full Scale IQ (RR = 1.12; CI 1.01, 1.24) and WRAT Spelling (RR = 1.14; CI 1.02, 1.28) and Arithmetic (RR = 1.15; CI 1.01, 1.31) tests.

Sensitivity analyses

Sensitivity analyses that took into account potential error in dating from LMP validated findings. The error-adjusted simulated GA weeks analysis demonstrated the same pattern of increasing average Bayley Mental and Psychomotor scores with each weekly increase in GA at delivery from 37 to 40 weeks, with an inflection point and improvement at 41 weeks. We observed a similar pattern of low scores, with notable results for WISC: FSIQ ($\beta = -1.78$; CI $-2.77, -0.79$) and all WRAT scores at 37 weeks (Table 5). There were no appreciable differences in estimates from analyses excluding women with gestational hypertension, pre-eclampsia or gestational diabetes.

Discussion

In a large, diverse US study with child follow-up through to age 7 years, children's neurocognitive development improved with each week of GA at delivery up to 40 or 41 weeks. These associations were strongest for 8-month Bayley assessments and 7-year WRAT results, but were also consistent for other assessments conducted at 4 and 7 years of age. The pattern of developmental improvement through to 40 weeks remained when evaluating the risk of scoring suspect/abnormal or below average on developmental tests and when evaluating within-family effects,

which accounted for unmeasured maternal or familial characteristics that would remain constant between siblings. The implications of our findings are that neurocognitive weekly gains continue into the term period, with optimal long-term neurocognitive outcomes conferred to children delivered past 39 weeks of gestation. Our findings are particularly salient in light of a recent practice advisory from the American College of Obstetricians and Gynecologists supporting elective induction at 39 weeks vs waiting for spontaneous labour (expectant management).¹⁷ This advisory was informed by the results of the ARRIVE trial, which found no differences in severe perinatal morbidity and mortality between women induced at 39 weeks and those who were expectantly managed up until 42 weeks 2 days (median delivery 40.0 weeks).²²

Our study is novel in that we evaluated each week of term individually, which is more clinically relevant than grouping weeks together given that clinicians make decisions about when to deliver pregnancies on a weekly basis.²³ Only three other studies have been conducted in the USA to assess neurodevelopment and all grouped delivery weeks at term together, limiting clinical application.^{1,3,9} In a recent study, Werner *et al.* evaluated differences in offspring school achievement of women who were induced at 39 or 40 weeks compared with those who were expectantly managed beyond those time periods and found no elevated risk for poor school performance at 8 years between groups.²⁴ Though our 39-week estimates lacked precision, we found a slight cognitive disadvantage for children born at 39 vs 40 weeks. This slight discrepancy may be explained by a few factors; in the Werner study, the reference group was heterogeneous, grouping outcomes across women who delivered any time after 39 weeks, 6 days instead of weekly comparisons as with our study. As observed in the results of ours and prior studies,^{12,25}

Table 5. Sensitivity analysis of the impact of error-adjusted GA weeks on the associations between gestational age at delivery and neurocognitive outcomes showing average mean difference in scores from reference (40 weeks) based on 1000 simulated error-adjusted replicates of GA weeks ($n = 39\ 199$)

	Gestational age at delivery (weeks)					
	37	38	39	40	41	42+
8-month						
Bayley Mental	-2.13 (-2.80, -1.41)	-1.41 (-2.09, -0.75)	-0.66 (-1.26, -0.04)	[Reference]	0.50 (-0.14, 1.13)	0.75 (0.22, 1.28)
Bayley Psychomotor	-3.31 (-4.11, -2.45)	-2.09 (-2.92, -1.28)	-0.94 (-1.70, -0.24)	[Reference]	0.64 (-0.11, 1.49)	1.01 (0.36, 1.67)
4-year						
SBIQ	-0.68 (-1.46, 0.11)	-0.34 (-1.08, 0.50)	-0.07 (-0.80, 0.71)	[Reference]	-0.04 (-0.77, 0.75)	0.08 (-0.59, 0.76)
7-year						
WISC: VIQ	-0.57 (-1.27, 0.12)	-0.34 (-1.01, 0.33)	-0.1 (-0.76, 0.51)	[Reference]	-0.01 (-0.70, 0.66)	-0.26 (-0.83, 0.3)
WISC: PIQ	-0.52 (-1.30, 0.29)	-0.27 (-1.02, 0.51)	-0.05 (-0.78, 0.68)	[Reference]	-0.10 (-0.83, 0.67)	-0.18 (-0.77, 0.43)
WISC: FSIQ	-0.75 (-1.47, -0.01)	-0.38 (-1.08, 0.26)	-0.11 (-0.81, 0.53)	[Reference]	-0.06 (-0.73, 0.63)	-0.25 (-0.84, 0.32)
WRAT: Spelling	-0.79 (-1.44, -0.13)	-0.36 (-1.00, 0.25)	-0.08 (-0.64, 0.54)	[Reference]	-0.13 (-0.77, 0.48)	-0.46 (-1.00, 0.03)
WRAT: Reading	-0.91 (-1.74, -0.14)	-0.47 (-1.33, 0.29)	-0.12 (-0.87, 0.65)	[Reference]	-0.11 (-0.88, 0.65)	-0.67 (-1.33, -0.03)
WRAT: Arithmetic	-0.59 (-1.20, -0.02)	-0.31 (-0.92, 0.28)	-0.07 (-0.62, 0.53)	[Reference]	-0.09 (-0.71, 0.42)	-0.43 (-0.94, 0.02)

Results of sensitivity analysis with generalized estimating equations for sibling clusters, adjusted for maternal age, socio-economic index, tobacco use, parity, pre-eclampsia, gestational hypertension, child race and study site, applied to each replicate of error-adjusted GA weeks, presented as β : average estimate based on 1000 replicates; 95% empirical confidence interval. FSIQ, Full Scale IQ; PIQ, performance IQ; SBIQ, Stanford-Binet Intelligence (IQ) test; VIQ, verbal IQ; WISC, Wechsler Intelligence Scales for Children; WRAT, Wide-Range Achievement Test.

developmental gains begin to decline in the post-term and consequently grouping women who delivered at 40+ weeks may mask scores that would be observed at each individual week of gestation. Furthermore, given the small effect sizes observed in our study, the study by Werner *et al.* may have been under-powered to detect an effect, even with a sample size of 6000. Moreover, the results may be confounded by not adjusting for the indication for induction, which may influence developmental outcomes. Taken together, our findings build upon previous studies that suggest that school performance, IQ and other cognitive outcomes may improve with increased length of gestation up to 40–41 weeks and decline post-term (42–45 weeks).^{3,4,9–16}

Our results are also consistent with the biologic evidence that brain development continues into the term period (37–41 weeks),^{2,26–28} where increases in grey matter and cerebral volume are observed through to 41 weeks.² Additionally, linear increases have been observed in neuroimaging studies in total grey matter,²⁷ temporal grey matter density,² cortical volume²⁸ and more efficient brain networks with each additional week of gestation.²⁶

One strength of our study was that we only included women who were admitted to the hospital in spontaneous labour. This inclusion reduced the likelihood of confounding by indication, such as pre-eclampsia and other pregnancy complications, which would be an indication for induction or scheduled caesarean delivery, and could also impact fetal brain development and later cognitive function.^{29,30} Additionally, the CPP is a rich source of data, with a large sample that is racial/ethnically diverse, has longitudinal follow-up and used reliable, validated assessments conducted by trained clinicians. Though conducted decades ago, all assessments utilized to define neurocognitive outcomes are in current use in updated formats. Additionally, it has been demonstrated that Bayley Mental tests conducted between 7 and 10 months were predictive of later cognitive outcomes³¹ and the original iteration of the Bayley Mental scale used in this study had high predictive power for long-term cognitive development.³² In addition to using data from multiple sites across the USA, we also used data from reliable and validated developmental assessments collected at multiple time points. Overall, our findings may translate to better generalization of results, more accurate interpretation and a broader picture of the progression of development than previous studies using measures of development assessed at a single point in time.^{3,4,9–12}

Assessment of GA based on LMP is a limitation of the study. However, our results were consistent in terms of magnitude and direction after attempting to account for errors across multiple sensitivity analyses, including robust

simulation scenarios. Additionally, the mean difference between LMP and ultrasound is generally 1–2 days,^{20,21} with the largest discrepancies associated with maternal demographic factors.³³ Our results were consistent across sibling analyses, where errors in LMP may be more consistent within children of the same mother than error across the full sample because of maternal demographic characteristics that may be consistent between pregnancies. However, despite this limitation, the main advantage of using the CPP is in terms of the size of the pregnancy cohort and length of follow-up time for offspring, which has continued into adulthood for some sites of original data collection.³⁴ Additionally, given the high rate of obstetric intervention in contemporary practice, it may be challenging to fully account for the indications of intervention in a modern study, which may contribute to both earlier GA at delivery (vs spontaneous labour) and long-term outcomes. To contextualize the importance of this point, in the CPP, the rates of caesarean delivery and induction of labour were 5.3% and 6.3%, respectively. Though reliable surveillance of obstetric intervention did not begin in the USA until the 1970s, the prevalence of interventions observed in our data is consistent with US estimates from the earliest years of reported data. Specifically, the rate of caesarean delivery in the USA was 5.5% in 1970³⁵ compared with 31.7% in 2019³⁶ and the rate of induction of labour was 7–9% in 1990^{37,38} compared with 29.4% in 2019.³⁶

We did not have information on alcohol use during pregnancy, which may be an important confounder. However, assuming consistent use between pregnancies, this factor should be accounted for in within-family analyses. Another limitation to our study is that some children were not assessed at all three time points, which could introduce attrition-related bias. However, the demographics of those available for all three assessments ($n = 24\ 363$) and those with missing data for any assessment were nearly identical (data not shown). We also utilized inverse-probability weighting to account for missing data at any given assessment point.

Conclusions

Neurocognitive development improved with every week of gestation beyond 37 weeks through to 40 or 41 weeks. Our findings, in conjunction with evidence that brain development continues with advancing gestation,^{26–28} highlight the importance of considering long-term outcomes in decision-making for non-medically indicated deliveries.

Supplementary data

Supplementary data are available at *IJE* online.

Ethics approval

This secondary analysis of existing data was exempt from human subjects' review.

Funding

This research was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

Data availability

All data, study forms and protocols are publicly available through the US National Archives (<https://www.archives.gov/research/electronic-records/nih.html>).

Author contributions

J.L.G. helped to conceptualize the project and completed data analysis in consultation with R.S. and A.S. J.L.G. drafted and revised all sections of the manuscript. S.E.G. and D.L.P. provided subject-matter expertise and consultation on the data. E.Y. and Y.V. helped to revise the manuscript and provided methodological input. K.L.G. conceptualized and supervised all aspects of the project. All authors reviewed and helped to revise the final manuscript.

Conflict of interest

None declared.

References

1. Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. *Pediatrics* 2009;123:e622–29.
2. Davis EP, Buss C, Muftuler LT *et al*. Children's brain development benefits from longer gestation. *Front Psychol* 2011;2:1.
3. Noble KG, Fifer WP, Rauh VA, Nomura Y, Andrews HF. Academic achievement varies with gestational age among children born at term. *Pediatrics* 2012;130:e257–64.
4. Searle AK, Smithers LG, Chittleborough CR, Gregory TA, Lynch JW. Gestational age and school achievement: a population study. *Arch Dis Child Fetal Neonatal Ed* 2017;102:F409–16.
5. Kirkegaard I, Obel C, Hedegaard M, Henriksen TB. Gestational age and birth weight in relation to school performance of 10-year-old children: a follow-up study of children born after 32 completed weeks. *Pediatrics* 2006;118:1600–06.
6. Ahlsson F, Kaijser M, Adami J, Lundgren M, Palme M. School performance after preterm birth. *Epidemiology* 2015;26:106–11.
7. Chan E, Quigley MA. School performance at age 7 years in late preterm and early term birth: a cohort study. *Arch Dis Child Fetal Neonatal Ed* 2014;99:F451–57.

8. Quigley MA, Poulsen G, Boyle E *et al.* Early term and late pre-term birth are associated with poorer school performance at age 5 years: a cohort study. *Arch Dis Child Fetal Neonatal Ed* 2012; **97**:F167–73.
9. Figlio DN, Guryan J, Karbownik K, Roth J. Long-term cognitive and health outcomes of school-aged children who were born late-term vs full-term. *JAMA Pediatr* 2016; **170**:758–64.
10. Smithers LG, Searle AK, Chittleborough CR, Scheil W, Brinkman SA, Lynch JW. A whole-of-population study of term and post-term gestational age at birth and children's development. *BJOG* 2015; **122**:1303–11.
11. Yang S, Bergvall N, Cnattingius S, Kramer MS. Gestational age differences in health and development among young Swedish men born at term. *Int J Epidemiol* 2010; **39**:1240–49.
12. Yang S, Platt RW, Kramer MS. Variation in child cognitive ability by week of gestation among healthy term births. *Am J Epidemiol* 2010; **171**:399–406.
13. Nielsen TM, Pedersen MV, Milidou I, Glavind J, Henriksen TB. Long-term cognition and behavior in children born at early term gestation: a systematic review. *Acta Obstet Gynecol Scand* 2019; **98**:1227–34.
14. Chan E, Leong P, Malouf R, Quigley MA. Long-term cognitive and school outcomes of late-preterm and early-term births: a systematic review. *Child Care Health Dev* 2016; **42**:297–312.
15. Hua J, Sun J, Cao Z *et al.* Differentiating the cognitive development of early-term births in infants and toddlers: a cross-sectional study in China. *BMJ Open* 2019; **9**:e025275.
16. Espel EV, Glynn LM, Sandman CA, Davis EP. Longer gestation among children born full term influences cognitive and motor development. *PLoS One* 2014; **9**:e113758.
17. American College of Obstetricians and Gynecologists. Practice Advisory: Clinical guidance for integration of the findings of The ARRIVE Trial: Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. 2018, Washington, DC: The College.
18. Klebanoff MA. The Collaborative Perinatal Project: a 50-year retrospective. *Paediatr Perinat Epidemiol* 2009; **23**:2–8.
19. Myrionthopoulos NC, French KS. An application of the U.S. Bureau of the Census socioeconomic index to a large, diversified patient population. *Soc Sci Med* 1968; **2**:283–99.
20. Hoffman CS, Messer LC, Mendola P, Savitz DA, Herring AH, Hartmann KE. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. *Paediatr Perinat Epidemiol* 2008; **22**:587–96.
21. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol* 2007; **21**(Suppl 2):86–96.
22. Grobman WA, Rice MM, Reddy UM *et al.*; Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal–Fetal Medicine Units Network. Labor induction versus expectant management in low-risk nulliparous women. *N Engl J Med* 2018; **379**:513–23.
23. Spong CY, Mercer BM, D'Alton M, Kilpatrick S, Blackwell S, Saade G. Timing of indicated late-preterm and early-term birth. *Obstet Gynecol* 2011; **118**:323–33.
24. Werner EF, Schlichting LE, Grobman WA, Viner-Brown S, Clark M, Vivier PM. Association of term labor induction vs expectant management with child academic outcomes. *JAMA Netw Open* 2020; **3**:e202503.
25. MacKay DF, Smith GCS, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 school children. *PLoS Med* 2010; **7**:e1000289.
26. Kim D-J, Davis EP, Sandman CA *et al.* Longer gestation is associated with more efficient brain networks in preadolescent children. *Neuroimage* 2014; **100**:619–27.
27. Huppi PS, Warfield S, Kikinis R *et al.* Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Ann Neurol* 1998; **43**:224–35.
28. Kinney HC. The near-term (late preterm) human brain and risk for periventricular leukomalacia: a review. *Semin Perinatol* 2006; **30**:81–88.
29. Gumusoglu SB, Chilukuri ASS, Santillan DA, Santillan MK, Stevens HE. Neurodevelopmental outcomes of prenatal preeclampsia exposure. *Trends Neurosci* 2020; **43**:253–68.
30. Camprubi Robles M, Campoy C, Garcia Fernandez L, Lopez-Pedrosa JM, Rueda R, Martin MJ. Maternal diabetes and cognitive performance in the offspring: a systematic review and meta-analysis. *PLoS One* 2015; **10**:e0142583.
31. Krogh MT, Væver MS, MS. A longitudinal study of the predictive validity of the Bayley-III scales and subtests. *Eur J Dev Psychol* 2019; **16**:727–38.
32. Ramey CT, Campbell FA, Nicholson JE. The predictive power of the Bayley Scales of Infant Development and the Stanford-Binet Intelligence Test in a relatively constant environment. *Child Dev* 1973; **44**:790–95.
33. Kullinger M, Westrom J, Kieler H, Skalkidou A. Maternal and fetal characteristics affect discrepancies between pregnancy-dating methods: a population-based cross-sectional register study. *Acta Obstet Gynecol Scand* 2017; **96**:86–95.
34. Buka SL, Shenassa ED, Niaura R. Elevated risk of tobacco dependence among offspring of mothers who smoked during pregnancy: a 30-year prospective study. *Am J Psychiatry* 2003; **160**:1978–84.
35. Placek PJ, Taffel SM. Trends in cesarean section rates for the United States, 1970–78. *Public Health Rep* 1980; **95**:540–48.
36. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK. Births: final data for 2019. *Natl Vital Stat Rep* 2021; **70**:1–51.
37. MacDorman MF, Mathews TJ, Martin JA, Malloy MH. Trends and characteristics of induced labour in the United States, 1989–98. *Paediatr Perinat Epidemiol* 2002; **16**:263–73.
38. Kozak LJ, Weeks JD. U.S. trends in obstetric procedures, 1990–2000. *Birth* 2002; **29**:157–61.