ASSESSING GESTATIONAL AGE AND SCHOOL PERFORMANCE AMONG SIBLINGS BORN IN NEW YORK CITY, 1994 - 1998

By

CARLA M. CLARK, MPH

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Public Health

Written under the direction of

Marian R. Passannante

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2020

ABSTRACT OF THE DISSERTATION

Assessing Gestational Age and School Performance among Siblings Born in New York City, 1994-1998

By CARLA M. CLARK

Dissertation Director:

Marian R. Passannante

Background: While many studies have established prematurity (births occurring before 37 weeks of gestation) as a major risk factor for neonatal morbidity and mortality, and poorer cognitive ability, the evidence on poorer cognitive outcomes/school performance in early term births (births occurring at 37-38 weeks gestation) versus that of full term births (39-41 weeks) is inconsistent. This dissertation seeks to add to the existing literature by assessing the relationship between gestational age (specifically early term births) and school performance controlling for the impact of unmeasured qualities of the family environment that may impact this association by using data from siblings.

Methods: This study uses linked birth registry and school records from the Longitudinal Study of Early Development (LSED) data warehouse, and includes children born in New York City between 1994 and 1998 who attended a New York City Public School during the 3rd grade. School performance is measured based on 3rd grade Math and English Language Arts (ELA) standardized test z-scores. Gestational age is based on the obstetric estimate from the birth certificate. Analysis utilizes a hybrid random effects model to allow for simultaneous estimation of within and between sibling group effects of gestational age on ELA and Math z-scores. Models

control for various demographic and clinical characteristics previously associated with gestational age and school outcomes.

Results: A total of 31,647 births in 15,432 sibling groups (range 2-4 siblings/group) were included in the analysis. Modeling revealed that a within sibling group difference in gestational age was not associated with a significant difference in ELA z-score (β=-0.003, 95% CI: -0.03, 0.03) nor Math z-score (β=0.007, 95% CI: -0.02, 0.04). However, several other variables (e.g. birth order, blood lead level, size for gestational age) were associated with significant within sibling group differences in ELA and Math z-scores.

Conclusions: Being born early term (37-38 weeks) versus full term (39-41 weeks) was not associated with standardized test score performance. Among children born at 37 or greater weeks' gestation, familial, social and environmental factors that have been shown to increase the probability of giving birth before or after 37 weeks are also associated with poorer school outcomes.

ACKNOWLEDGEMENTS

I begin by giving thanks for the blessings from above that allowed me to complete this journey. I am deeply indebted to my advisor, Dr. Marian Passannante, whose PATIENCE, encouragement and understanding buoyed me through this saga. Many thanks to Dr. Soyeon Kim, whose sage advice and guidance helped immensely throughout the data analysis phase. Deep gratitude also to Dr. Kitaw Demissie an early supporter of my doctoral education and whose encouragement and advice I am grateful for. Thank you to Dr. Rahjeev Mehta for being part of this committee and for taking the time out to participate in conference calls and impart your clinical insight on this process.

To my dear current and former colleagues at the New York City Department of Health and Mental Hygiene, thank you for the encouragement and for being great thought partners as I began and continued this work.

Thank you to my parents, brothers, sisters-in-law, nieces, and nephews, aunts, uncles and cousins for your love, support and quiet cheerleading.

A very special thanks to my husband; you held my hand and believed in me when I didn't think I could finish this. I dedicate this dissertation to my sons. Know that the sky's the limit; always believe, and you can achieve.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Gestational age at birth plays a fundamental part in determining infant health and survival.(1) Several studies over the last several decades have established prematurity (births delivered before 37 completed weeks' gestation) as a major risk factor for neonatal morbidity and mortality, and poorer long-term outcomes including cognitive ability at school age and early-age school performance compared to full-term infants (37-41 weeks' gestation). (2–10) This finding is further supported by reports describing increased brain development with increasing gestational age when comparing preterm to term infants. (11–13)

More recently, births occurring at 37-38 weeks' gestation known as 'early term' births (14,15) have been shown to be associated with poorer neonatal outcomes. (1,16) In 2009 the rate of infant mortality for live births delivered at 37-38 weeks was 56% higher than for those delivered at 39-41 weeks' gestation (3.09/1,000 births vs. 1.98/1,000 births, respectively). (1) This disparity has persisted over time.(17) Evidence of the impact of early term birth on long-term outcomes is mixed. (18) Some studies of the impact of early term birth on long-term outcomes, including a meta-analyses have reported an increased risk of lower IQ scores, higher special educational need, behavioral, emotional and neurodevelopmental issues, as well as poorer subject-matter standardized test scores at every gestational age in the term period below 39 weeks. (9,10,19–29) However, other studies show no significant relationship. (6,18) Given that over 25% of births in the U.S. occur in the early term. (1,30,31), differences in long term outcomes among children born early term could have a significant population impact.

While the studies of the association between gestational age and school outcomes control for many of the maternal and socio-demographic characteristics that may confound this relationship, they are limited in other important respects. One primary limitation is that many within-family characteristics that may influence the association between gestational age and

school outcomes, such as parental IQ and family home environment remain inadequately or completely unmeasured and are uncontrolled for in most analyses. Another concern is that the maternal conditions and pregnancy complications which may vary across pregnancies are not adequately measured or controlled for in some analyses. These conditions may be an underlying cause of early birth and a source of harm to fetal brain development. (25,32,33) Finally, of the studies which assess the relationship between term birth and school outcomes among cohorts which include multiple births, only a few more recent studies show evidence of taking into account the correlation between siblings included in analyses; correlated data do not meet the assumption of independence required for linear regression and some other multivariable models. (4,34–36) The studies which do account for the correlated nature of sibling data often estimate the sibling and whole-population relationships separately. Other considerations include the age of the cohorts (important given the advances in medical technology over time (24)) and how both gestational age and school performance are measured.

In their article on racial and ethnic disparities in birth outcomes, Lu and Halfon (37), propose combining two mechanisms which have been used to explain the influence of a woman's health across the life span on future pregnancy outcomes: early programming (38) and cumulative pathways (38–41) to examine racial/ethnic disparities in birth outcomes. In the conceptual schematic developed by Lu and Halfon (37) shown in Figure 1.1 , the horizontal axis represents the time periods in the life span and the vertical axis, a woman's reproductive potential (e.g. cognitive capacity of the child), and it shows that there is a cumulative and dynamic effect of risk factors over time that play a role in affecting women's health and reproductive outcomes. The authors posit that there is a persistently lower likelihood for positive reproductive outcomes for African American women, relative to White women resulting from the earlier (even beginning in utero) and more frequent exposure to risk factors (represented by the downward arrows) without the buffer of protective factors (the upward arrows) so their outlook for positive pregnancy outcomes is diminished with time and age.

To further illustrate, Figure 1.2 shows the infant mortality rate by race/ethnicity and maternal education in NYC in 2011. It is well known that education is an important driver of health such that one would expect racial differences to diminish significantly once outcomes are stratified by education; they do not in this example. Infants born to non-Hispanic, Black women in NYC experience higher rates of infant mortality than any other racial/ethnic group regardless of maternal education. Moreover, the infant mortality rates for infants born to college-educated Black women are higher than for infants born to mothers of other races who never graduated high school suggesting that the effect of race surpasses that of educational attainment. Data in the NYC Pregnancy Risk Assessment Monitoring System (PRAMS) show that many black college educated women who give birth in NYC (compared to white college educated women) were nearly two and a half times as likely to have more financial and other stressors/ adverse life events which have occurred continuously over their lives as well as during pregnancy, including earning less money for similar work as other groups, often serving as the sole breadwinner, and being less likely to have insurance (NYC PRAMS unpublished data). (42) The differential in incomes by race despite comparable education has been documented elsewhere.(43) Applying the concepts put forth by Lu and Halfon, the continuous influence of these risk factors without the presence of protective factors to buffer their effects contribute to an accelerated decline in health such that their pregnancy outcomes are poorer versus women from other groups of the same educational status. (37) The observation that some racial/ethnic groups are not able to realize the benefits of higher SES has been documented by others as well. (44,45) Given the evidence above, and the complexity of relationships between various factors which may not even be measured, or are measured poorly, this dissertation intends to assess whether the relationship between gestational age in the term period and school outcomes varies by race/ethnicity. To the extent that in utero exposures (e.g. maternal stress, maternal chronic conditions etc.) affect cognitive development (35,46–49), the increased benefits from longer stay in utero reported in many studies (i.e. delivering at later gestational age) may not be realized for some groups; even among

small-for-gestational age children who experience catch-up growth, there may not be a corresponding development in cognitive ability, which impacts school outcomes. (50) Additionally, to the extent that various stressors continue into the post-partum period and beyond without the buffer of protective factors, and occur differentially by race/ethnicity, the environment needed to support continued cognitive development of the child is compromised for some groups. (51,52) Therefore, it is important to assess risks and outcomes at the subgroup level as they remain important in understanding the nature of the relationship between gestational and school performance.

Figure 1.1. Differential exposures resulting in poorer outcomes. Source: Lu and Halfon, 2003.

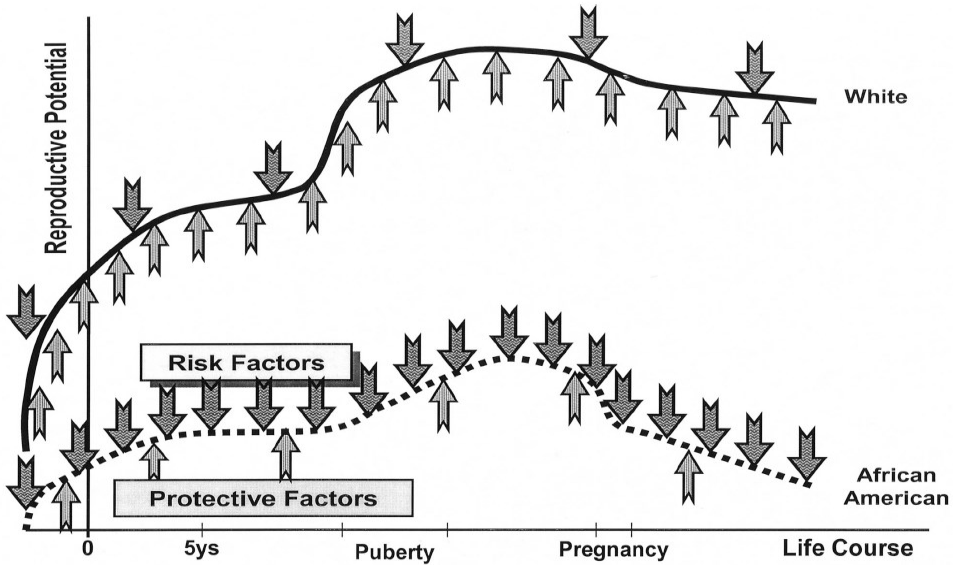
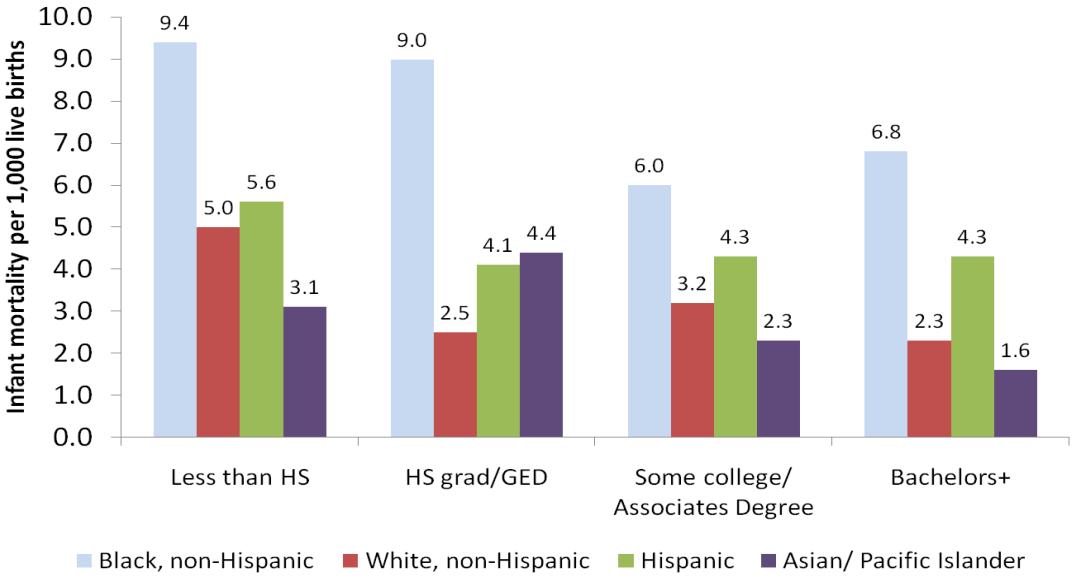


Figure 2.2. Infant Mortality by Race/Ethnicity and Maternal Education, NYC 2011 Source: NYCDOHMH. Office of Vital Statistics, compiled by Bureau of Maternal, Infant and Reproductive Health



*note: HS, high school; GED, general education development.

This dissertation seeks to answer the following:

Research Question 1: Is there a difference in standardized test score performance, by gestational age among siblings born in the term period in New York City?

Hypothesis 1: There is a difference in school performance on standardized tests among siblings born in the early term vs full term period.

Research Question 2: Does the relationship between gestational age and school performance vary according to race/ethnicity?

Hypothesis 2: The relationship between gestational age and school performance among children born in the term period varies according to race/ethnicity

This analysis will contribute to the growing research on early term births and scholastic achievement and seeks to improve on the previous analyses conducted in NYC by simultaneously assessing the relationship between early term gestation and school performance within and between sibling groups. Using sibling groups establishes control for the impact of unmeasured qualities of the family environment that can influence such associations. The analysis will also examine whether variation in this relationship exists by race/ethnicity, a consideration not explored previously using NYC data; it will also take into account the role of pregnancy and obstetric complications on long term scholastic outcomes. The analysis will use propensity scoring to control for confounding by various maternal complications and risk factors while reducing the degrees of freedom in the model.

CHAPTER 2

LITERATURE REVIEW

Summary

This chapter aims to summarize the existing literature exploring the relationship between gestational age and school outcomes. In the summary below, studies on gestational age and school outcomes are divided into general population based or non-sibling studies and sibling-based analyses. The primary reason for this dichotomy is to underscore the potential utility of sibling-based analyses in attaining the 'counterfactual' comparison group, by reducing the problem of uncontrolled confounding due to familial factors, which often remain inadequately or completely uncontrolled for in general population based studies.

General population (non-sibling) based studies

There are few studies of term gestational age and school outcomes using design and analytical techniques to control for unobserved within-family characteristics which play an important role in school performance. (4,34,53,54) Nonetheless, non-sibling based studies provide important insight as to the potential differences in development that may exist among a gestational age group (term births) that was once thought to be homogeneous in outcomes.

Two such studies assess the relationship between gestational age and school outcomes in New York City children. (10,26) While the primary outcome of the study conducted by Lipkind et al was school performance for births within the preterm period (early preterm (32-33 weeks) versus late preterm (34-36 weeks)) and compared to term births as a whole (37-42 weeks), the authors also assessed week-to-week differences in standardized testing scores across the entire gestational age spectrum (32-42 weeks); they found significant increases in test scores (~.1%-.8% of standard deviation) for each 1- week increase in gestational age, including in the term period, that persisted after controlling for a host of important demographic, obstetric, and

maternal medical risk factors that could potentially confound such a relationship. (10) In educational literature, the findings translate into small but non-negligible differences and improvements in test scores. (55) Nevertheless, some important limitations remained. Firstly, there was no evidence of taking into account the correlation between siblings, despite the presence of siblings in the study, violating of the assumption of independence required for logistic regression and other multivariable models used in the study. This was of particular importance given the study cohort consisted of births from multiple years (1994-1998), among whom it is estimated that approximately 20% had at least 1 sibling. (56) In addition, factors such as alcohol use, and lead exposure which are known to affect brain function (49) were also not controlled for; depending on the prevalence of these factors in the population, estimates of effect could be biased; for example in 1995, nearly 20,000 children aged 6 months to 5 years of age were identified with elevated blood lead levels (BLL)(>10 μg/dL) in New York City. (57) Other factors such as family home environment, and parental IQ (which may not always be reflected by educational attainment), can influence school performance (58) but were not controlled for in the analysis by Lipkind et al. Noble et al. focused exclusively on school performance among children born in the term period in a cohort of children born between 1988 and 1992. (26) Of interest, unadjusted, but significant improvements in reading and math scores by weekly gestational age grouping were found up until the 39th week of gestation. However, the authors did not present a model which considered all significant obstetric, individual, and community level factors in tandem to assess the relationship between gestational age in the term period and school performance (measured by 3rd grade standardized scores); rather, they assessed whether and found that gestational age remained significant in 3 separate models controlling for obstetric, individual level factors (e.g. race and age), and community-level factors respectively. Moreover, several factors remained uncontrolled for in each model. Specifically, in the model containing obstetric and medical factors, conditions such as hypertension, diabetes, and pre-eclampsia

remained uncontrolled for. These factors may be a cause for early birth and in the case of diabetes may also be implicated in reduced intellectual ability. (35)

Other non-sibling based studies of school outcomes and term gestational age have found higher adjusted risks and odds for poor school outcomes with lower gestational age in the term period. (9,21–23,27,29) These studies varied however in the length of follow-up (range: 5-19 years), levels of comparison made with respect to gestational age: continuous (week-to-week) versus categorical (37-38 vs. 39-41), and techniques to measure school outcomes including parental and teacher assessment of school achievement overall and for specific subjects, (21,23) special education need (27) , not completing basic school (9) , or post-secondary education (22), and performance in language literacy (29). Parental and teacher assessment of achievement in particular carries the risk of information bias as there was no inter-rater reliability or criterion validity of the assessments. Findings based on follow-up occurring in the later years of life are potentially affected by the myriad of events in the life course since birth that remain uncontrolled for in these analysis (e.g. parental loss, loss of income, childhood home environment, social/peer influences, survivor effects). Finally, other limitations include many of those found in the New York City studies described above (10,26), with respect to confounding control and statistical techniques.

Studies of gestational age and cognitive development have also found improvements in cognitive ability with increasing gestational age in the term period. (19,25) Specifically, Yang et al. found lower IQ scores for children age 6.5 years at each week decrement of gestational age at birth in the term period below 39 weeks after controlling for various maternal and family variables, including parity, behaviors during pregnancy (i.e. smoking, alcohol use), marital status, parental education and occupation, and maternal age. (25) However, it is not clear if the IQ point differential is meaningful with regards to the children's future school outcomes, and there is the potential that associations observed are the result of residual confounding due to lack of control

for some maternal health complications such as diabetes. Further, unmeasured characteristics of the family (cognitive stimulation, etc.) also remained unaccounted for.

An additional study of note was that conducted by Nepomnyashcy et al. (6) While the objective of the study was to measure the developmental outcomes of late-preterm infants at 2 and 4 years, the authors also assessed outcomes for early term vs. full term deliveries and found that there were no differences in developmental outcomes at 2 years of age after adjusting for demographic, obstetrical and economic covariates; only differences in vocabulary scores at age 4 years remained marginally significantly related to birth in early term vs. full term. Notably, including only demographic and economic covariates in a multivariable model did not produce estimates of the association between late-preterm birth and developmental outcomes substantially different than a model including maternal conditions and pregnancy complications alongside demographic and economic factors, suggesting that the relationship between gestational age at birth and school outcomes may not have been greatly influenced by clinical/obstetric factors in the population under study. Alternatively, underreporting of clinical information could also produce these results.

A study by Gurka et al. (59) examined the relationship between gestational age and several outcomes including scholastic achievement, cognitive development and emotional and behavioral. Interestingly, this study did not identify significant differences in outcomes between late-preterm and full term children. In addition to being based on a small sample (n=53 for late preterm births), study participants were volunteers, introducing the likelihood of selection bias, as those who agreed to participate may have had less neonatal and post-neonatal morbidity.

Chan and Quigley in a non-sibling, population-based study investigated the effect of late preterm (34-36) and early term (37-38 week) births on school performance in reading, writing and mathematics at age 7 years using the UK Millennium Cohort (14). Notably, while children born early term had statistically significant lower performance in subject-specific domains, they did not differ from full term (39-41 week) births on the primary outcome of achieving the

expected level of general performance in reading, writing and mathematics. This suggests perhaps the existence of an effect of gestational age on particular functioning (e.g. comprehension, speaking and listening). Missing from consideration in this as well as many other studies outlined here are the considerations of other family environmental factors (e.g. reading to children, help with homework) as well as considering differences in outcomes controlling for school/classroom level characteristics.

Smithers and colleagues in Australia used a local developmental index, the Australian Early Development Index (AEDI) to assess the level of development vulnerability for all gestational ages in the term range below 40 weeks, as well as births at 41 weeks and 42-45 weeks. (60) They observed births occurring at 40-41 week gestation to have the lowest risk of overall developmental vulnerability at school entry while those born at 37, 38 and 39 weeks had 13%, 5% and 2%, respectively, increased (though non-statistically significant) risk for vulnerability on 1 or more AEDI domains. Notable here again as in the Chan and Quigley study, is that on several individual domains such as communication skills and general knowledge, births occurring at 37 weeks gestation had a non-statistically significant increased adjusted risk (aRR:1.23, 95%CI:0.96-1.58) of developmental vulnerability, while on language and cognitive skills there was a statistically significant increased adjusted risk of developmental vulnerability for births at 37 weeks compared to births at 40 weeks (aRR:1.34, 95%CI:1.03-1.74).

In 2015, a systematic review of the existing literature on this topic (reviewing many of the articles described herein) revealed mixed evidence on the association between late preterm and early term births and cognitive/ school performance. (18) In some, the adverse cognitive and school performance findings associated with preterm and early term birth attenuated and became non-significant once adjusted for risk factors, while others still found significant though attenuated differences.

Since then, a population-based study conducted by Searle and Smithers et al in 2017, again using data from Australia (n=28,155) examined whether achievement varies within the term

period (37-41 weeks). (3) With every week increase in gestational age within the term, they observed lower risk of poor 3rd grade numeracy scores and lower risk of poor performance in selected literacy domains. For example, compared to children born at 40 weeks, children born at 37 weeks had higher adjusted risk of being below standards for grammar (aRR=1.11, 95% CI:1.01,1.21) and numeracy (aRR=1.10, 95% CI:1.01,1.20). For reading, writing and spelling, elevated, but non-significant risks for poor performance were observed (Reading: aRR=1.09, 95% CI:0.99, 1.19; Writing: aRR=1.13, 95% CI:0.99,1.29; Spelling: aRR=1.04, 95% CI:0.95,1.14). This may be reflective of 'noise' in the data; however, the observation of variation in risk depending on function measured across studies in of note.

In a large Danish study (n=615,789), Wilngreen et al found that odds of needing special education support and failing to complete compulsory education increased over the range of gestational age at birth, using 40 weeks gestation as the reference. (2) This study also identified lower birthweight for gestational age standardized z-scores (-2 standard deviation score (SDS)), to be associated with a 96% higher odds of needing special education support, and a 71% higher odds of failing to complete compulsory school. Since the study included all live-born infants born in Denmark from 1992 to 1997, allowing for correlations among sibling pairs was indicated; however, this was not taken into consideration. Consequently, it is possible that results are overestimates of the true relationship between gestational age and these outcomes.

Another population-based study in Denmark examined gestational age and socioeconomic achievements in young adulthood. (61) In this study of 288,030 non-anomalous singleton births born in Denmark between 1982 and 1986, gestational age < 39 weeks was associated with lower odds of high educational level and high income. While these findings prove interesting, given that outcomes were measured at age 28, the impact of the post-natal environment, including family, social and physical environmental exposures are important considerations that arguably have a stronger influence on very long-term outcomes.

Sibling based studies

Since family environment is an important determinant of school performance (53,54,62) , controlling for family characteristics (educational encouragement in the home) would attenuate the differences in school outcomes by gestational age if they are due to family environment. Only 1 study has been identified that used a fixed family effects design to simultaneously measure within sibling and between non-sibling effects. The study noted no difference in birth weight for gestational age z-score and intelligence (measured via IQ score) within sibling pairs; however, the study found a robust increase in intelligence per week increase of non-birth weight standardized gestational age within siblings, which was slightly attenuated when compared with differences found in the whole cohort of participants (siblings and only children). (28) However, the study failed to control for several important confounders which can vary across pregnancies, including maternal smoking, and drinking during pregnancy.

Other sibling-based studies of birth characteristics and school outcomes examine the relationship between birth weight for gestational age (using a standardized z-score) and cognitive ability (which influences school performance). The z-score is often difficult to interpret or apply practically and may mask the importance of gestational age alone as an important predictor of school performance. (63–66) In addition, sibling-based studies in this area have yielded mixed results. Two studies (63,65) conducted both sibling and 'whole cohort' analyses and found no within-sibling differences in cognitive ability in relation to birth weight for gestational age z-score. Of note, in the study conducted by Yang et al (63), differences between birth weights were small and may not have translated into differences in gestational age. For example, the mean difference in birth weights in 2-sibling families was approximately 1 ounce; this may mean that these siblings did not differ in gestational age, and may account for the findings of no difference in sibling pairs. In addition, this study relied on maternal report of both birth weight and gestational age which is subject to recall bias and may have resulted in over or under estimation of these measures for one or more siblings.

Studies by Eriksen et al (66) and Matte et al (64) found significant positive within-family (sibling) differences in intelligence scores in relationship to z-scores of birth weight standardized to gestational age. Of note, Eriksen's study included males only, and did not control for factors that may vary across pregnancy, though findings were robust among siblings with large and smaller age differences. It was not clear however, whether the gestational age-specific means and standard deviations used to create z-scores were also gender standardized. If the general population on which the standardization was based has a higher proportion of female than male births, there is the potential for misclassification bias and inaccurate estimates of association. Matte et al also found evidence of an association between birth weight and IQ score for boy sibling pairs, but not girls. Interestingly the sample studied by Matte (a siblings sample drawn from the National Collaborative Perinatal Project database) had a higher than average SES (based on income, education, and occupation of head of household), was more likely to have white mothers, and were more likely to be born to mothers under 20 years of age, as compared to the general population, thus its generalizability is limited. In addition, while the study examined and found that sex modified the relationship between IQ and birth weight for gestational age, they did not examine this for other variables such as maternal age, which may have yielded other important findings.

Abel et al. in a population-based study of Swedish children found that end of secondary education grade averages of infants born post-term (>40 weeks) and preterm (<40 weeks) were lower than those of children born at 40 weeks. (4) In addition and importantly, lower averages were also observed for those born small for gestational age, regardless of gestational age (in the 1987-1994 birth years: -0.13 SD; 95%CI: -0.14, -0.12). This study also controlled for maternal and paternal psychiatric history; a consideration not assessed in other studies. Additional key features of this study are the inclusion of multiple birth cohorts across time (1973-1994), and in order to account for correlation of outcomes among siblings across birth cohorts, sibling pairs were matched preterm to term and term to post term. Though slightly attenuated, the associations

remained significant within sibling pairs strengthening the potential for a causal effect of gestational age on school performance. However, the sibling pairs analysis compared the sibling-averaged gestational age which is a weighted average of the within and between sibling effects, and thus is possibly an overestimation the true within-family effect.

Most recently, Heuvelman and colleagues examined the risk of intellectual disability across the gestational age spectrum, among youth in Stockholm with no evidence for disability linked to genetic or other inherited metabolic syndromes. (34) Intellectual disability was measured using a composite of information from several health care registries, including inpatient and outpatient diagnoses based on ICD-10 and DSM-IV indications. Using both whole cohort and matched sibling pairs, the study found odds of intellectual disability highest for those born extremely preterm (24 weeks), but also found 50%, 26%, and 10% higher odds of intellectual disability among 37, 38, and 39 week births, respectively, compared to those born at 40 weeks. These associations though attenuated, persisted in the sibling pairs analysis. Most notably, the researchers also found that infants born small for gestational age had increased odds of intellectual disability across the gestational age spectrum; that is whether preterm, term or post-term, children born small for gestational age had the highest odds of intellectual disability compared to those born appropriate for gestational age.

Many of the sibling and non-sibling based studies described previously also fail to or are inadequately powered to assess the potential interactive effects of variables such as race/ethnicity, gender and parental education on the relationship between gestational age and school outcomes. Among those which have assessed interaction, Yang et al (25) found no evidence of interaction with gender nor race in their study of gestational age in the term period and childhood cognitive ability. A single study was identified which examines parental education as an effect modifier (29), and found that the relationship between preterm birth and language performance varied according to parental education; a poorer performance in Swedish language class was only observed among preterm births where both parents had a lower educational level. This supports

the evidence that brain development continues well into the post-natal period, thus allowing for cognition and school performance to be influenced by external /post-natal factors. (67)

In both sibling and non-sibling based studies, the variability in findings may be due in part to differences in data collection methods, the populations under study, variables used/available to control for confounding, and the measurement of the outcome (intelligence/cognitive ability, school performance) and exposure (gestational age based on last menstrual period versus ultrasound dating).

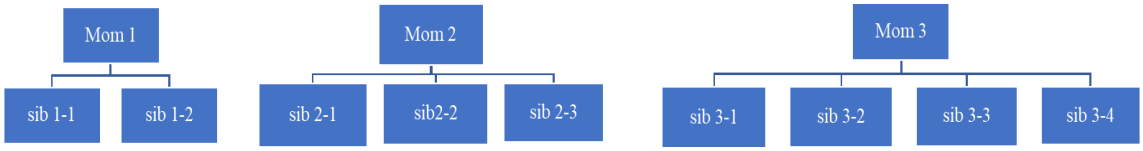
CHAPTER 3

METHODOLOGY

Study Design

This is a retrospective cohort study of singleton siblings born in New York City (NYC) to NYC resident mothers between 1994 and 1998 who were delivered at term gestation (37-41 weeks). Exposure is gestational age at delivery and the outcomes are grade 3 Math and English Language Arts (ELA) standardized test scores. Analysis utilizes the nested nature of this dataset, whereby children (level 1) born to the same mother are nested within sibling groups (level 2) [Figure 3.1]. Comparing siblings while accounting for the correlated nature of the data helps minimize residual confounding due to unobserved characteristics of the family environment that may remain uncontrolled for in population-based studies using analyses that ignore the clustered nature of the data. Interaction between race/ethnicity and gestational age is also examined to understand if there are variations in effect across groups.

Figure 3.1 Siblings nested within groups



Participant inclusion criteria:

Eligibility for inclusion in this study is based on criteria similar to that developed by Lipkind et. al (10), who were the first to utilize this dataset in the exploration of gestational age and school outcomes. Births were eligible for inclusion if they were singleton, term births in NYC to NYC resident mothers between 1994 and 1998, attended a NYC Public School during the

3rd grade and took the ELA and Math standardized tests, and who also had a sibling born at term gestation in NYC to the same mother between 1994 and 1998.

Participant exclusion criteria:

All children born in NYC who were missing gestational age, are only children (no siblings) or having siblings but not in the cohort under study, were born pre or post-term (<37 weeks or >41 weeks, respectively), born with birth anomalies known to affect cognitive development, who moved before school enrollment, were enrolled in private school or home schooled, who were missing one or both test scores (Math and/or ELA) or who attended a NYC Public School but were born outside of NYC, are excluded from the study sample. It is important to note that while some birth anomalies result in fatality soon after birth (e.g. trisomy 18) not all birth anomalies are fatal or result in impaired cognition. Children with birth anomalies who die before grade 3 (and thus have no test scores) are excluded; children who survive with these anomalies will also be excluded. Further, births with the following anomalies found to be associated with impaired cognition are excluded: all neural tube defects, chromosomal anomalies (e.g. trisomy 18, trisomy 21), cardiac congenital anomalies (specifically, cyanotic congenital heart disease), hydrocephalus, microcephalus, tracheo-esophageal fistula/atresia, omphalocele/gastroschisis, congenital rubella syndrome, encephalocele and renal anomalies. Information on the type of congenital anomaly is available from the birth record.

Data Source

The data analyzed in this study are from the Longitudinal Study of Early Development (LSED), a comprehensive data warehouse that links individualized information on children born between 1994 and 2004 and have at least 1 record in any of the 5 data sources included the LSED warehouse: NYC birth and death registries, the NYCDOH Lead Poisoning Prevention Program (LPPP) registry and housing inspection databases, administrative and financial databases of the Early Intervention Program, the NYC Department of Education administrative and special education databases and PLUTO files (geographic and census data).

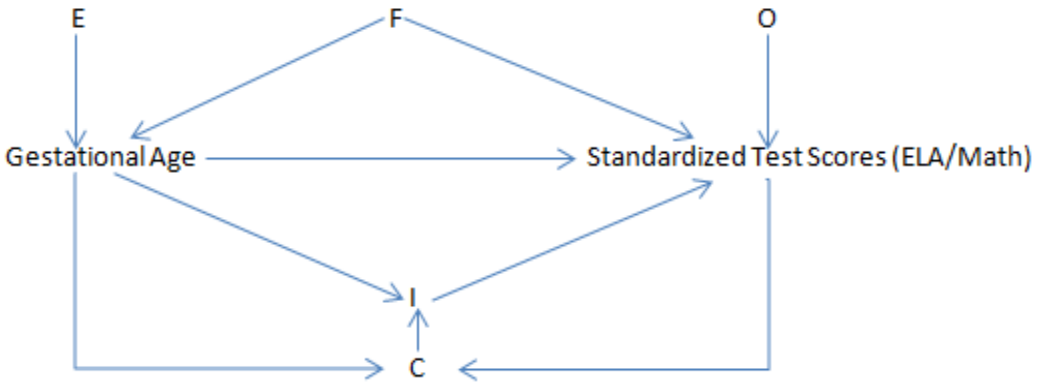
Data were linked using IBM's QualityStage 8.0 probabilistic linkage software and linkage/matching was assessed to have an accuracy range of 98.6-99.4%. The complete warehouse contains 1,942,942 uniquely identified children and 306,240 sibling sets. (56) The population of interest includes children born in NYC to NYC resident mothers between 1994 and 1998, and who were attending a NYC Public School during the 3rd grade. The Department of Education 3rd grade standardized testing data is available only for children born through 1998 (tested in 2006), because obtaining and matching the data was conducted by the NYCDOH in 2007; therefore, children born 1999 and later do not have test scores. School performance is measured based on 3rd grade Math and English Language Arts (ELA) standardized test scores, and gestational age is based on the obstetric gestational estimate from the birth certificate. This estimate is primarily based on first trimester ultrasound. The dataset is unique in many respects including the diversity of individuals and the sheer size of the sample, the availability of a broad spectrum of obstetrical and maternal data as well as child outcome data from several registries across NYC Health Department programs. (58,68)

Study Variables

A listing of study variables can be found in Appendix A (Table A1). Variables used in this study derive from 3 main sources within the LSED data warehouse: birth registry data, NYC Department of Education databases and Lead Poisoning Prevention Program registries. Demographic and clinical/obstetric/pregnancy specific study variables within the LSED data warehouse come from the NYC birth data which is generated from a worksheet completed by the parent at the time of birth, and the Confidential Medical Report of birth completed using the hospital and prenatal records for all births in NYC (see Appendix A, Figure A1). Standardized test (i.e. ELA and Math) scores are derived from the NYC Department of Education data in LSED, and information on the highest BLL from the LPPP registry data in LSED. 'Siblingid' is a variable created by the LSED warehouse to identify sibling groups (i.e. shared birth mother).

All covariates used to control for confounding were selected a priori based on findings in the literature. A directed acyclic graph (DAG), shown below, is used to illustrate the conceptual relationship between covariates and the exposure and outcomes in this study (Figure 3.2). Controlling for collider variables (i.e. variables which are an effect of both the exposure and the outcome), may result in biasing estimates in the opposite direction of the true effect. Further, controlling for variables along the causal pathway between exposure and outcomes (i.e. over-adjustment) may result in underestimation of the causal effect. (69–72)

Figure 3.2 Directed Acyclic Graph (DAG) for the effects of gestational age on 3rd Grade Standardized Test Scores.



Confounders (F)	**Predictors of outcome (O)**
propensity score variables	Highest Blood lead level
Birth order	Received school lunch[1]
Birth year (proxy for pregnancy interval)	Child sex
Race/ethnicity	Delivery method
Nativity	Apgar score
	Small for gestational age
	Neonatal Intensive Care Unit (NICU)

[1]Proxy for SES; the colliders (C) and intermediates (I) introduce bias into the estimated effect so adjustment should be avoided (70)
Predictors of exposure (E)

Exposure

The exposure, gestational age in the term period, is defined as births occurring between 37 weeks, 0 days and 41 weeks, 6 days. Births occurring between 37 weeks, 0 days and 38 weeks, 6 days are further defined as 'early term', and births between 39 weeks, 0 days and 41 weeks, 6 days defined as 'full term'. The dichotomization of the term period into early term and full term stems from several findings including the American College of Obstetricians and Gynecologists (ACOG) Committee Opinion which encouraged the use of this re-definition of term pregnancies in light of the differential neonatal outcomes observed from deliveries in these 2 gestational time frames (15). In the NYC birth certificate data base, gestational age is a discrete variable reported in number of completed weeks, and is based on obstetric gestational estimate derived from first trimester ultrasound dating; the date of the last menstrual period (LMP) is used when ultrasound data are unavailable.

Outcomes

The outcomes of interest are 3rd grade Math and ELA standardized test scores (administered to this 1994-1998 birth cohort between 2002 and 2006 respectively), measured on a continuous scale. The tests are administered to NYC students by the NYC Department of Education and attempt to measure a student's knowledge base and skill set in the 3rd grade subjects of Math and ELA. Math and ELA test used from 2003 to 2005 were citywide, while those administered between 2006 and 2007 were statewide; As a result, Math and ELA test scores will be transformed to z-scores by child's year of birth to standardize the scale on which the tests are measured; this was also done in the analysis conducted by Lipkind et al. (10) Z-scores were calculated by subtracting the mean test score for a given year of birth from an individual's test score and dividing this by the standard deviation of birth year-specific test score. (10,66) Unlike IQ tests, subject-specific standardized tests are more widely administered in school settings to children at various grade levels; thus the potential applicability of differences found in

standardized test scores by gestational age group may be more widespread than differences in IQ scores which are reported in some studies.

Covariates

Demographic variables included mother's self-reported race/ethnicity, measured as black, non-Latino; white, non-Latino; Latino; Asian/Pacific Islander or Other. For women who had discordant race/ethnicity identification across birth records, the race of the mother indicated on the first birth with non-missing/non-unknown race is used as the mother's race/ethnicity. Mother's and father's age measured as age at last birthday prior to child's birth are continuous variables, though for preliminary descriptions of the population of interest they are reported in the following categories: <20 years old, 20-34 years old, 35+ years. Mother's and father's education completed included the categories: less than high school; high school; some college; 4 years college or more. Mother's nativity is a categorical variable based on self-reported country of origin. US born includes birth in the continental US, Hawaii, Puerto Rico, Guam, or other US territory. Birth in all other countries is considered non US born. Mother's employment during the pregnancy as well as whether the child was ever eligible for school lunch program are both binary (yes/no) variables. Primary financial coverage for pregnancy and birth, i.e. insurance status is a 4-category variable that includes: Health Maintenance Organization (HMO), other 3rd party, self-pay, and Medicaid. Mother's weight is based on the pre-pregnancy weight as listed on the confidential medical report and included in the birth file data from the LSED warehouse. Because height was not routinely collected on the medical report during the study period, body-mass index (BMI) cannot be calculated. Thus, based on pre-pregnancy values, weight is categorized as obese, overweight, or normal. Diabetes is measured as chronic diabetes; gestational diabetes; no diabetes. Pregnancy-related hypertension includes eclampsia; preeclampsia; pregnancy associated hypertension or none of these conditions. Method of delivery is summarized as vaginal or cesarean section (c-section). Child's year of birth refers to the years included in this study: 1994, 1995, 1996, 1997, and 1998. This is also a crude proxy for birth interval. Birth order refers to the

order in which each birth has occurred within a family (1st, 2nd, 3rd, etc.). In the LSED warehouse, multiple births share a number. Child's sex assigned at birth is binary: male or female. Apgar score at 5 minutes, which assesses how well the neonate is doing ex utero, is measured on a scale of 1 to 10. Any congenital anomaly, coded as yes/no, measures if any of the births in the cohort have one of the included categories of congenital anomalies not believed to impact cognitive outcomes. Highest venous blood-lead level test is reported in micrograms per deciliter (μg/dl). In descriptive tables it is presented categorically (0-4μg/dl, 5-10 μg/dl, >10 μg/dl), but is used as a continuous variable in comparative analysis. The following pregnancy-related, clinical and obstetric variables are coded binary yes/no: mother's tobacco use during pregnancy; mother's alcohol use during pregnancy; mother's use of amphetamines, cocaine, heroin, methadone or marijuana; chronic hypertension; no prenatal care received; neonatal intensive care unit admission; medical risk factor - previous preterm or small for gestational age infant; medical risk factor - renal disease; other maternal risk factors (includes anemia, acute or chronic lung disease, uterine bleeding, incompetent cervix, STD presence, hydramnios/oligohydramnios); complication of labor/delivery - cord prolapse; complication of labor/delivery - fetal distress; complication of labor/delivery - abruptio placenta; complication of labor/delivery - placenta previa; small for gestational age (measured as <10th percentile of birthweight for gestational age). Finally, marital status, while available in the data warehouse was excluded from this analysis due to irregularities in the determination of the number of marital and non-marital births. Specifically, an algorithm in use at the time these data were collected for the birth registry was overestimating the number of non-marital births; if the misclassification was differential, this could bias estimates. (73)

Analyses

Data analysis includes 1- calculating the frequency and distribution of exposure, outcomes, and covariates; 2- assessing the relationships between covariates and the exposure and outcomes; 3- multiple imputation for covariates with missing data; 4- estimation of propensity scores and 5- estimation of the relationship between standardized test scores and gestational age.

Data in clusters such as sibling groups are often correlated. The relatedness of observations violates the assumption of independence needed for employing techniques such as Ordinary Least Squares (OLS) regression for the estimation of effects (74–80) . In addition, when the clustered nature of the data is ignored and techniques for non-clustered data are employed, standard errors are underestimated calling into question the validity of differences in observed estimates. (74–77,79–83) Multilevel models help to address this challenge by accounting for the variation occurring at both the individual and cluster level. Variables included in this study are measured at level 1 (individual sibling), or at level 2 (sibling group). Level 1 variables are partitioned into their within and between sibling group components in order to describe for each variable, the amount of variation in the outcome that is accounted for by differences within the sibling group (within effect) and the amount accounted for by differences between sibling groups (between effect). This is particularly important in the context of this study as it seeks to understand the role gestational age in the term period plays in affecting standardized test scores. A finding of no relationship between term gestational age and standardized test scores within sibling groups may suggest that differences in the family environment for different sibling groups is responsible for the variation observed in standardized test scores. All data analysis is conducted using SAS® Version 9.4 (SAS Institute Inc., Cary, North Carolina, USA).

1994-1998 birth cohorts

To better understand the population of interest, frequencies and percentages for the demographic, clinical/obstetric/pregnancy and school-related (i.e. test score) characteristics of the 1994- 1998 birth cohorts are calculated and the distribution of test scores are examined using box plots. Next, the characteristics of included and excluded births of 37-41 weeks gestation are compared using a Generalized Estimating Equation (GEE) with exchangeable correlation structure.

Descriptive and comparative analysis of exposure, outcomes and covariates

Comparative estimates of the distribution of demographic and clinical characteristics of births by exposure (gestational age) and outcomes (ELA and Math z-scores) are calculated using random effects models which account for variability in the outcome across sibling groups. Estimates are calculated using the variables in their non-partitioned (original) form, thus providing a weighted average of within and between sibling group variations in estimation. PROC GLIMMIX is used for estimating relationships with gestational age and PROC MIXED is used in assessing relationships to ELA and Math z-scores.

Multiple Imputation for missing covariates

Analysis that uses only those observations with complete information on all variables can lead to biased estimates of effect if there is some differential distribution in completeness (84–89). Multiple imputation techniques are frequently employed to handle missing data on the exposure, the outcome, other predictors used in the analysis, or any combination therein. Compared to some other techniques for handling or 'filling in' missing data, multiple imputation infuses a level of variation around the true value since the observed (non-missing) data are used to estimate multiple values (89). Important to the process of multiple imputation is understanding what missing data mechanism is present in the data set. In this study sample, the missing data are assumed to follow a missing at random (MAR) mechanism, whereby other variables in the dataset can be used to predict missingness for a variable (but the variable itself cannot be used to predict its own missingness). (84,88,89) There are 3 main phases in multiple imputations: the imputation phase, the analysis phase, and the pooling phase. (84,88,89) Included below is a summary of how each phase is implemented in this study.

(1) Imputation Phase

Single-level multiple imputation procedures were applied to the data set for this study, ignoring the multilevel structure. In simulation studies, when values were missing for covariates and the multilevel nature of the data set was not accounted for, the true value regression coefficients for the variables was either over or underestimated depending on the magnitude of

the intraclass correlation coefficient (ICC). (90,91) Consultation with local experts in multiple imputation suggested that for this specific analysis, single level multiple imputation may be feasible given the size of the clusters (personal communication, Dr. Jason Roy, 12/2018). In addition, census tract poverty level, and percent of persons African-American in census tract are 2 community level auxiliary variables used in the imputation models. Auxiliary variables have been noted to improve the quality of the imputed values generated. (89) Further, the use of community level variables with single-level multiple imputation of multilevel data has been shown to be a good proxy for having a random intercept (i.e. taking clustering into account). (92)

Because several variable types with missing values exist in the data set, the Fully Conditional Specification (FCS) approach was selected to allow separate imputations for each variable with missing values using different distributions for each imputed variable. (89) Discriminant function or logistic regression was used for categorical variables with missing data, and linear regression for continuous variables. Research suggests that imputation models should be 'congenial' with the final model and thus contain variables in such form as they appear in the final model (89); however, it has also been shown that using the FCS method with the variables partitioned into their within and between components could introduce bias if the mean being subtracted is skewed by missing values. (93,94) Therefore, the imputations were carried out on the variables in their non-partitioned form. Also recommended in the literature, the number of imputations should equal to the highest fraction of missing information (FMI) percentage as this affects results the most. (89) The FMI which is based on the percent missing for a variable and its correlation with other variables in the imputation model was assessed and the number of imputations selected therein. Twenty burn-in iterations were done such that an imputed data set was saved after every 20^{th} computational cycle.

(2) Analysis Phase

The multivariate analysis of data in this study (described in subsequent sections) involves calculation of a propensity score, and utilizes propensity score information in subsequent

multilevel models of the outcome. Propensity scores are calculated for each observation in each of the imputed data sets then used in the related imputation for the substantive model. Importantly, analysis was first conducted on a single imputed dataset, and additional data decisions made before proceeding with estimation across other imputed data sets.

(3) Pooling Phase

In the final phase of the imputation process, estimates from each imputed and analyzed data set are pooled to describe outcomes.

Propensity score calculation

The outcome of performance on standardized tests is being compared among gestational ages in the term period. Many maternal risk factors and comorbidities have been shown to lead to early birth. As previously mentioned these factors may also be a source of harm to fetal brain development and thus affect school outcomes. In this study where multiple covariates need to be controlled, propensity scores provide a way to adjust for the numerous covariates with fewer degrees of freedom. Estimated propensity scores can be used in a number of ways to control for confounding: matching, stratification, inverse probability of treatment weight and covariate adjustment (95–98). These methods can be used singularly or in tandem if indicated (95). For this study, the probability of being born at 39-41 weeks was calculated for each birth using a hybrid random-effects model to compute both within and between sibling group estimates. While fixed-effects modeling using conditional logistic regression model (99) has typically been used when clustered data are involved, fixed effects models cannot calculate the effects of level 2 (i.e. between sibling group) variables on the outcome. Model variables included selected demographic, clinical/obstetric and pregnancy covariates (shown in Appendix A, Table A1) that a priori are known to be associated with gestational age and school test scores as well as those strongly related to school test scores (even if unrelated to gestational age). Brookhart et al. has recommended this strategy as way to reduce variance and mean square errors in estimates. (100) The form of the regression model used to calculate the propensity scores is as follows:

P(gestational age) = {Between and within sibling components for group of observed confounders}

where {group of observed confounders} = maternal comorbidities and pregnancy complications, maternal pregnancy behaviors, and selected correlates of school test scores partitioned into their within and between cluster components (see next section for explanation of cluster components). Following estimation, predicted values (propensity scores) from the model were stratified into ranks based on quintiles (95) and those quintiles included in the outcome models partitioned into their within and between components, rather than matching on the propensity score which would reduce the sample size. (101) Of note however, the use of propensity scores for covariate adjustment does not handle the issue of unbalanced data unless the values are restricted to the range of propensity scores common to both the 37-38 week (early term) and 39-41 week (full term) births which may also bias estimates if the region of overlap is small. (101) PROC GLIMMIX is used to estimate the propensity score for each imputed data set. The GLIMMIX procedure fits statistical models known as generalized linear mixed models (GLMM), for data with correlations (such as in this study) or non-constant variability and where the response may not be normally distributed. (102)

Estimation of the relationship between gestational age and standardized test scores

The major questions explored using these data are:

Is there a difference in standardized test score performance, by gestational age, among siblings born in the term period in NYC? and *Does the relationship between gestational age and school performance vary according to race/ethnicity?*

The hypotheses are (respectively):

1. There is a difference in school performance on standardized tests among siblings born in the early term vs full term period.

2. The relationship between gestational age and school performance among children born in the term period varies according to race/ethnicity.

The relationship between gestational age and school performance in the whole cohorts (1994-1998) has been assessed previously. (10) In the current study, the relationship between gestational age and school performance is assessed among siblings in the cohort of study eligible singletons born at 37 to 41 weeks gestation between 1994 and 1998 using a hybrid random effects model. Model building strategies also integrated the assessment of whether the relationship between gestational age and school performance among siblings varies according to race/ethnicity and nativity. Variables (not included in the propensity score) that are a priori known to be associated with math and ELA scores and with gestational age are considered potential confounders and are included in multivariable analyses. A hybrid random-effects model was computed using PROC MIXED, to estimate the relationship between gestational age, and Math and ELA z-scores concurrently within and between sibling groups (63,82,103). Simultaneous estimation of within sibling and between sibling group effects allows for assessment of whether fixed family factors or pregnancy specific/intrauterine factors may explain associations between gestational age and school outcomes. (82) For each model, the formula is as follows (63,74,75,79,80) :

$$E(Y_{ij}) = \beta_0 + \beta_w(X_{ij} - X_j') + \beta_B X_j' + \beta_1 (X_{1ij} - X_{1j}') + \ldots + \beta_k (X_{kij} - X_{kj}') + \beta_{1B} X_{1j}' + \ldots + \beta_{kB} X_{kj}' + (u_j + e_{ij})$$

Where $E(Y_{ij})$ is the expected Math or ELA z-score for a given sibling i in sibling group j; X_j' is the mean gestational age for each sibling group (used for between group effect); X_{ij} is the gestational age for an individual sibling i in sibling group j; and $X_{ij} - X_j'$ is the deviation of the individual sibling gestational age from the sibling group mean gestational age (used in calculating the within-sibling effect). β_w is the within-sibling group coefficient that measures the change in Math or ELA z-score for a unit change in the deviation of an individual sibling's gestational age from the sibling group gestational age. β_B represents the between sibling groups coefficient that

measures the change in average Math and ELA z-score for every unit change in X_j'. Covariates (e.g. propensity score ranks) for each sibling *i* in family *j* are represented by X_{1ij} *thru* X_{kij} and coefficients=β_1 *thru* β_k (for factors varying within siblings) respectively; and for each family *j* the mean of covariates X_{1j}' *thru* X_{kj}' represent factors varying across (between) sibling groups. Finally, u_j and e_{ij} represent the error terms at the sibling group and individual sibling level, respectively.

The model building process is summarized in Table 3.1. Model building included estimating a model with no covariates (Model 0), followed by adding in the predictor (gestational age); then remaining level 1 variables. A random slope model, used to assess whether the slope of the within sibling group gestational age varies across sibling groups was then estimated. Next, level 2 variables (Race/ethnicity, nativity) were added to the model. Least Squares estimates were computed to test for differences in the linear combinations of race-ethnicity, nativity. To test for interaction, the final model tested included an additional parameter to assess the interaction of race/ethnicity and gestational age between sibling groups:

$$E(Y_{ij}) = \beta_0 + \beta_w(X_{ij} - X_j') + \beta_B X_j' + \beta_1 (X_{1ij} - X_{1j}') + \ldots + \beta_k (X_{kij} - X_{kj}') + \beta_{1B} X_{1j}' + \ldots + \beta_{kB} X_{kj}' + \underline{\beta_z X_j' X_{kj}'} + (u_j + e_{ij})$$

where βz = coefficient for interaction of race and gestational age between sibling groups; and X_j' X_{kj}' =the value for the interaction of race/ethnicity and mean gestational age for a given sibling group.

Table 3.1 Summary of Model Building for Propensity Score and ELA and Math z-Scores

Model	**Propensity (confounder) Score**	**ELA and Math z-score**
Model 0	No covariates; includes only random effect for the intercept (level 2 variance).	No covariates; includes only random effect for the intercept (level 2 variance).
Model 1	All level-1 fixed effects partitioned into their within and between cluster components	Model 0 + gestational age group partitioned into its within and between cluster components

Model	Propensity (confounder) Score	ELA and Math z-score
Model 2	---	Model 1 + remaining level-1 fixed effects partitioned into their within and between cluster components
Model 3	---	Model 2 + random slope for gestational age group
Model 4	---	Model 3 + level-2 fixed effect (Race/ethnicity-nativity composite)
Model 5	---	Model 4 + terms to test for Interaction between Race/ethnicity-nativity and gestational age group

Following model estimation for each imputed data set, PROC MI Analyze is used to compute summary estimates (β, 95% confidence Intervals (CI) and standard errors) over the imputed data sets, and Akaike Information Criterion (AIC) is used to compare the goodness of fit for the models computed. Regression diagnostics were computed and outliers explored using residual plots.

Sensitivity analysis

To test the robustness of the final model selected and examine stability of estimates, 3 sensitivity analyses were conducted using the final model selected. This first was an analysis of the original dataset which includes missing data on covariates (complete case analysis). Next a model was computed using only those births to women who were first time mothers at age 40 or older to increase the likelihood that estimates are being made on an entire sibling group, and not just a sample as this may also lead to biased estimates. Lastly, a fixed effects model is computed, which estimates only the within sibling group effect of gestational age on the test scores. Thus, the impact of race/ethnicity and nativity cannot be computed for this model. All models are computed using a significance level of 0.05.

CHAPTER 4

RESULTS

Study Sample Description

Between 1994 and 1998, there was a total of 973,763 births with at least 1 record in any of the data sources contributing to the LSED data warehouse. Of these, 589,666 births occurred in NYC to NYC resident women. The characteristics of these births are summarized in Table 4.1. Slightly under 10% of births were preterm (<37 weeks gestation), and 20.89% were early term (37-38 weeks). The majority of births occurred at full term (39-41 weeks); while nearly two and a half percent were born post term (42-44 weeks).More than 60% of births occurred to Black, non-Latino and Latino women. Women were on average nearly 28 years old at their last birthday prior to giving birth [mean (sd):27.69 (6.31)], and slightly over one-third had completed high school (36.35%). An almost even number of births to US born and non US born women occurred between 1994 and 1998 (50.35% versus 49.06%, respectively). For fathers in the sample, while there was a significant amount of information missing on age (23.83% missing) and education (26.18% missing), those with data were on average 31 years old at their last birthday prior to the child's birth [mean (sd):31.47 (7.16)], and slightly under one-third had completed high school (30.84%). Census tract poverty level, a measure of neighborhood poverty, showed that a slightly higher proportion of births (28.73%) occurred to residents of very high poverty neighborhoods (>= 30% of residents in the census tract live below 100% of the federal poverty level) in New York City. A proxy individual measure of poverty, child eligibility for the school lunch program shows that approximately 37% of children ever having attended a NYC public school met this definition. Notably this metric is missing for more than half of births (53.43%) occurring between 1994 and 1998, as many children either did not attend a NYC Public School at any time during the 3rd grade or earlier, or perhaps did not provide information to make this determination. Most mothers were not employed during their pregnancies (61.43%), and more than half of all births

were paid for by Medicaid (54.51%). Over three-quarters of women's pre-pregnancy weight were reported as normal (77.48%). Reported tobacco use during pregnancy was 5.13%. There was a very low frequency of drug use (including amphetamines, cocaine, heroin, methadone or marijuana) and alcohol use reported (1.33% and 0.5%, respectively). Among clinical characteristics, chronic diabetes was very infrequent (0.26%) and gestational diabetes occurred at 3.64%. Overall, pregnancy related hypertension (including eclampsia, preeclampsia and pregnancy-associated hypertension) occurred for nearly 4% of births. Chronic hypertension was <1% and slightly over 2% of pregnancies resulting in a live birth received no prenatal care. Among the medical risk factor present among mothers for this population of births, most occurred at <5%. Having had a previous preterm (<37 weeks) or small for gestation infant occurred at 0.38% and renal disease at 0.10%. Other maternal risk factors (as seen in Chapter 3) occurred at 12.21%. Most births were delivered vaginally (76.46%). Among the complications occurring at labor/delivery fetal distress occurred for 3.09% of births in this population, while cord prolapse, placenta previa and abruptio placenta occurred at 0.15%, 0.23%, and 0.29% respectively. Births occurred approximately equally across the years of this cohort: 1994 births (20.98%); 1995 (20.54%); 1996 (19.87%); 1997 (19.25%); 1998 (19.37%). Most infants were first in the birth order (87.31%), though this number is slightly inflated as it includes multiple births, who share a birth order number. Slightly over half of births were assigned male sex at birth (51.14%) and 48.86% assigned female at birth. Five-minute Apgar scores were overwhelmingly 7 or higher (98.36%), and <1% of births had a congenital anomaly not considered impactful to cognition. Approximately 14% of births were small for gestational age, and 8.22% were admitted to the NICU. Finally, the highest venous blood-lead level test was <5 mg/dl for most births (33.83%).

Table 4.1 Characteristics of Births in New York City (NYC) to NYC Resident Mothers, 1994-1998 (n=589,666)

Characteristic	Frequency	Percent
Mother's Race/Ethnicity		
black, non-Latino	172727	29.29
white, non-Latino	150078	25.45
Latino	204255	34.64
Asian/Pacific Islander	57566	9.76
Other	2914	0.49
Unknown/Missing	2126	0.36
Mother's Age		
< 20 years old	63453	10.76
20-34 years old	435045	73.78
35+ years old	90975	15.43
Missing	193	0.03
mean (sd)	27.69	(6.31)
Mother's Education Completed		
Less than high school	152345	25.84
High School	214373	36.35
Some College	103586	17.57
4 years College or more	101451	17.20
Missing	17911	3.04
Mother's Nativity		
US born	296919	50.35
nonUS born	289303	49.06
Missing	3444	0.58
Father's Age		
< 20 years old	13907	2.36
20-34 years old	293575	49.79
35+ years old	141669	24.03
Missing	140515	23.83
mean (sd)	31.47	(7.16)
Father's Education Completed		
Less than high school	79562	13.49
High School	181880	30.84
Some College	73167	12.41
4 years College or more	100686	17.08
Missing	154371	26.18

Characteristic	Frequency	Percent
Census Tract poverty level*		
Low poverty (<10%)	143967	24.42
Medium poverty (10% to <20%)	147208	24.96
High poverty (20% to <30%)	127937	21.70
Very High poverty (>= 30%)	169432	28.73
Unknown/missing	1122	0.19
Was child ever eligible for school lunch program**		
Yes	219709	37.26
No	54903	9.31
Missing	315054	53.43
Mother Employed During This Pregnancy		
Yes	167158	28.35
No	362230	61.43
Missing	60278	10.22
Primary Financial Coverage for Pregnancy and Birth		
HMO	62961	10.68
Other 3rd Party	161717	27.43
Self-pay	31126	5.28
Medicaid	321427	54.51
Missing	12435	2.11
Mother's Weight***		
Normal weight	456888	77.48
Overweight	29960	5.08
Obese	964	0.16
Missing	101854	17.27
Mother's tobacco use during pregnancy		
Tobacco use	30274	5.13
No tobacco use	550663	93.39
Missing	8729	1.48
Mother's Alcohol Use During Pregnancy		
Yes	2956	0.5
No	573773	97.3
Missing	12937	2.19
Mother's Use of Amphetamines, cocaine, heroin, methadone or marijuana		

Characteristic	Frequency	Percent
Yes	7869	1.33
No	581797	98.67
Diabetes		
Chronic diabetes	1562	0.26
Gestational diabetes	21440	3.64
No diabetes	566664	96.1
Pregnancy-related hypertension		
Eclampsia	418	0.07
Preeclampsia	13890	2.36
Pregnancy-associated hypertension	6043	1.02
No pregnancy-related hypertension	569315	96.55
Chronic hypertension		
Yes	5105	0.87
No	584561	99.13
No prenatal care received		
Yes	13205	2.24
No	531698	90.17
Missing	44763	7.59
Medical Risk Factor - Previous Preterm or SGA Infant		
Yes	2241	0.38
No	500228	84.83
Missing	87197	14.79
Medical Risk Factor - Renal Disease		
Yes	570	0.1
No	501899	85.12
Missing	87197	14.79
Other maternal risk factors		
At least 1 misc. maternal risk factor	72002	12.21
None of the misc. Maternal risk factors	517664	87.79
Complication of Labor/Delivery - Cord Prolapse		
Yes	874	0.15
No	501595	85.06
Missing	87197	14.79
Complication of Labor/Delivery - Fetal Distress		

Characteristic	Frequency	Percent
Yes	18245	3.09
No	484224	82.12
Missing	87197	14.79
Complication of Labor/Delivery - Abruption Placenta		
Yes	1692	0.29
No	500777	84.93
Missing	87197	14.79
Complication of Labor/Delivery - Placenta Previa		
Yes	1342	0.23
No	501127	84.98
Missing	87197	14.79
Method of Delivery		
Vaginal	450849	76.46
C-section	128897	21.86
Unknown/Missing	9920	1.68
Child's year of birth		
1994	123687	20.98
1995	121097	20.54
1996	117187	19.87
1997	113494	19.25
1998	114201	19.37
Birth Order (multiple births share a number)		
1st	514813	87.31
2nd	68982	11.7
3rd or higher	5871	0.99
Child's Sex assigned a birth		
Female	288092	48.86
Male	301574	51.14
Apgar Score, 5 Minutes		
<7	6092	1.03
7+	579976	98.36
Unknown/Missing	3598	0.61
Any Congenital Anomaly		
Yes	4110	0.7
No	585556	99.3

Characteristic	Frequency	Percent
Small for Gestational Age (<10th percentile of birthweight for gestational age)		
Yes	83925	14.23
No	502174	85.16
Missing	3567	0.6
Neonatal Intensive Care Unit (NICU) Admission		
Yes	48486	8.22
No	425357	72.14
Missing	115823	19.64
Gestational Age at Delivery		
22-36 weeks	55634	9.43
37-38 weeks	123166	20.89
39-41 weeks	389786	66.1
42-44 weeks	14683	2.49
Unknown/Missing/Invalid	6397	1.08
Highest venous blood-lead level test		
0-4µg/dl	199455	33.83
5-10 µg/dl	100090	16.97
>10 µg/dl	12018	2.04
Missing	278103	47.16
mean (sd)	4.54	(3.49)

*% of residents living below 100% of the federal poverty level

**a proxy measure of individual level poverty

***Only weight information was available on the medical report accompanying birth certificate submissions; therefore, Body Mass Index (BMI) could not be calculated.

Table 4.2 displays the frequency of grade 3 ELA and math standardized test scores by proficiency level. Notably most of the births in this population were not enrolled in a NYC public school or had unknown or missing grade 3 proficiency (63.56% for ELA and 61.38% for math). Among those with scores, most fell into the 2nd or 3rd level for grade 3 ELA proficiency (26.92%); while most scored at the 3rd or 4th level of proficiency for grade 3 Math (26.99%). Box plots and an accompanying summary of the z-score distribution for 3rd grade ELA and Math test scores by gestational age are presented in figures 4.1 and 4.2, respectively. Both ELA and Math

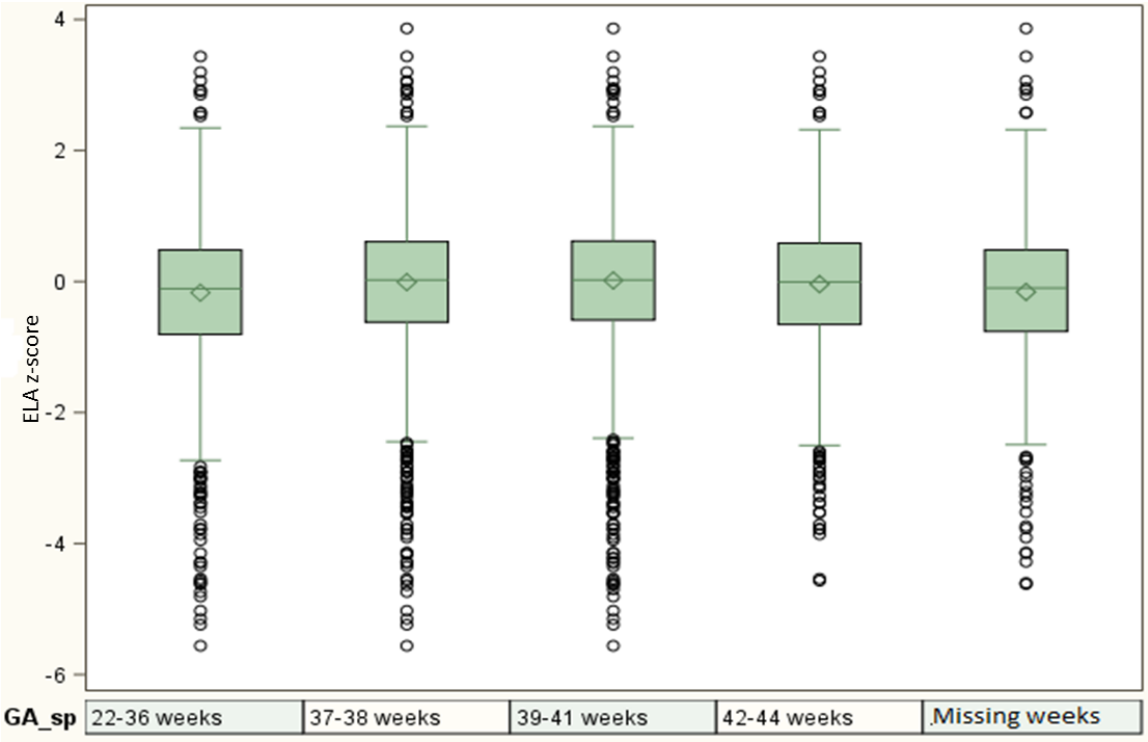
z-scores appear similarly distributed with similar means for early term and full term births, while preterm (more notably) and post-term births have lower means.

Table 4.2 Earliest Grade 3 Standardized Test Scores of Children born in New York City (NYC) to NYC Resident Mothers 1994-1998 (n=589,666)

Standardized Test Scores[a]	**Frequency**	**Percent**
Grade 3 ELA proficiency		
Level 1	30586	5.19
Level 2	66654	11.30
Level 3	92089	15.62
Level 4	25573	4.34
Unknown/Missing/Not in DOE	374764	63.56
Grade 3 Math proficiency		
Level 1	23561	4.00
Level 2	44969	7.63
Level 3	94670	16.05
Level 4	64524	10.94
Unknown/Missing/Not in DOE	361942	61.38

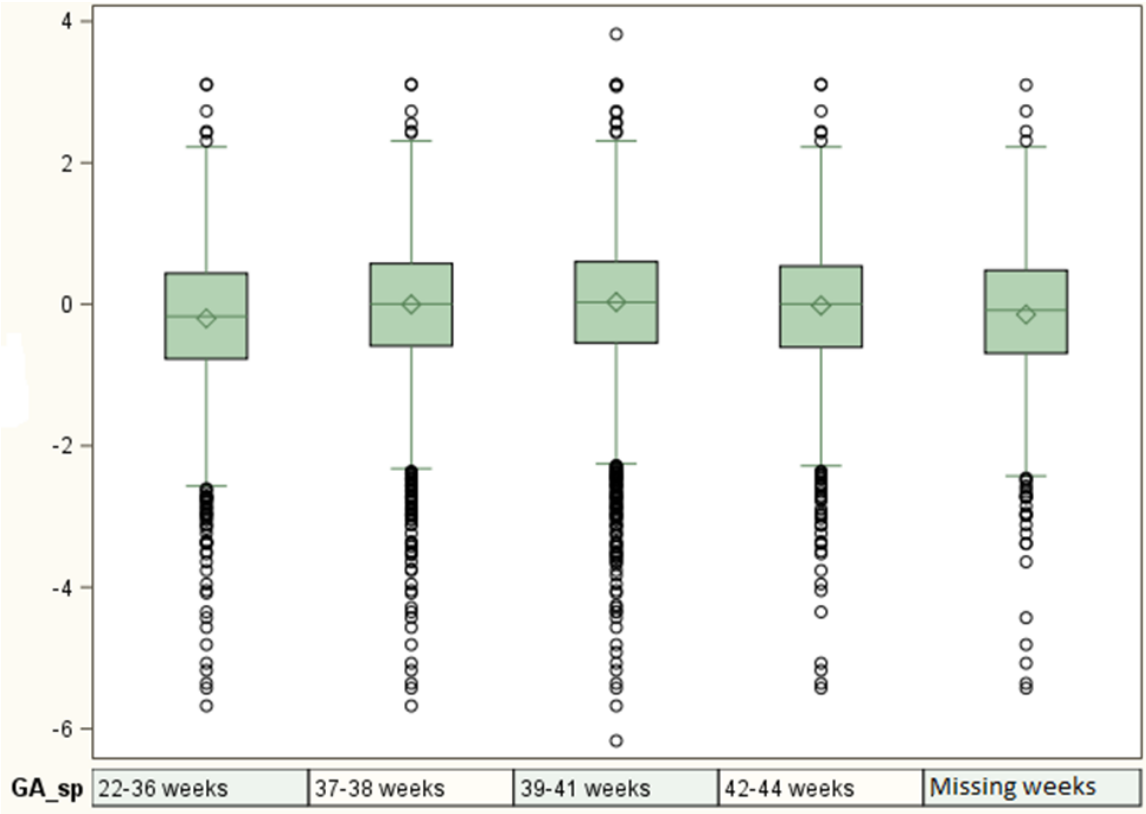
[a] ELA and Math standardized tests were administered by New York City through 2005 and by New York State thereafter; as a result the numerical scores included in each group are not mutually exclusive. Proficiency levels are labeled above with 1 representing having scored well below standards and 4 representing above standards.

Figure 4.1 Relationship between gestational age group and ELA Standardized Test Score (z-scores) among NYC births to NYC mothers, 1994-1998 (n=214, 902)



Gestational Age Group	22-36 weeks	37-38 weeks	39-41 weeks	42-44 weeks	Missing Weeks
Total N	55634	123166	389786	14683	6397
Valid n	20820	46531	140116	5149	2286
n missing	34814	76635	249670	9534	4111
Mean	-0.17	-0.01	0.02	-0.04	-0.16
Median	-0.11	0.02	0.02	-0.01	-0.1
25th%ile	-0.81	-0.62	-0.59	-0.65	-0.76
75th%ile	0.48	0.61	0.62	0.59	0.48
Minimum	-5.56	-5.56	-5.56	-4.57	-4.62
Maximum	3.44	3.87	3.87	3.44	3.87

Figure 4.2 Relationship between gestational age group and Grade 3 Math Standardized Test Score (z-scores) among NYC births to NYC mothers, 1994-1998 (n=227, 724)



Gestational Age Group	22-36 weeks	37-38 weeks	39-41 weeks	42-44 weeks	Missing Weeks
Total N	55634	123166	389786	14683	6397
Valid n	21881	49170	148789	5475	2409
n missing	33753	73996	240997	9208	3988
Mean	-0.2	0	0.03	-0.02	-0.14
Median	-0.17	0	0.03	0	-0.08
25th%ile	-0.77	-0.59	-0.55	-0.61	-0.69
75th%ile	0.44	0.58	0.6	0.54	0.48
Minimum	-5.67	-5.67	-6.17	-5.43	-5.43
Maximum	3.11	3.11	3.82	3.11	3.1

Final Study Sample

The following categories of births were excluded: gestational age that was either missing (n=6,397), pre-term(n=55,634), or post-term (n=14,683); non-singleton births of 37-41 weeks gestation (n=7,926); congenital anomalies impacting cognitive development (n=839); not in the NYC Department of Education databases (i.e. attended private, charter, or home school) (n=270,304); missing Math and/or ELA test score (n=50,549); no sibling in the 1994-1998 cohort or is an only child (n=138,184); eligible siblings of births excluded earlier due to ineligibility (n=13,503). This resulted in a final analytic sample consisting of 31,647 births in 15,432 sibling groups of 2-4 children (Fig 4.3).

Figure 4.3 Sample selection flowchart

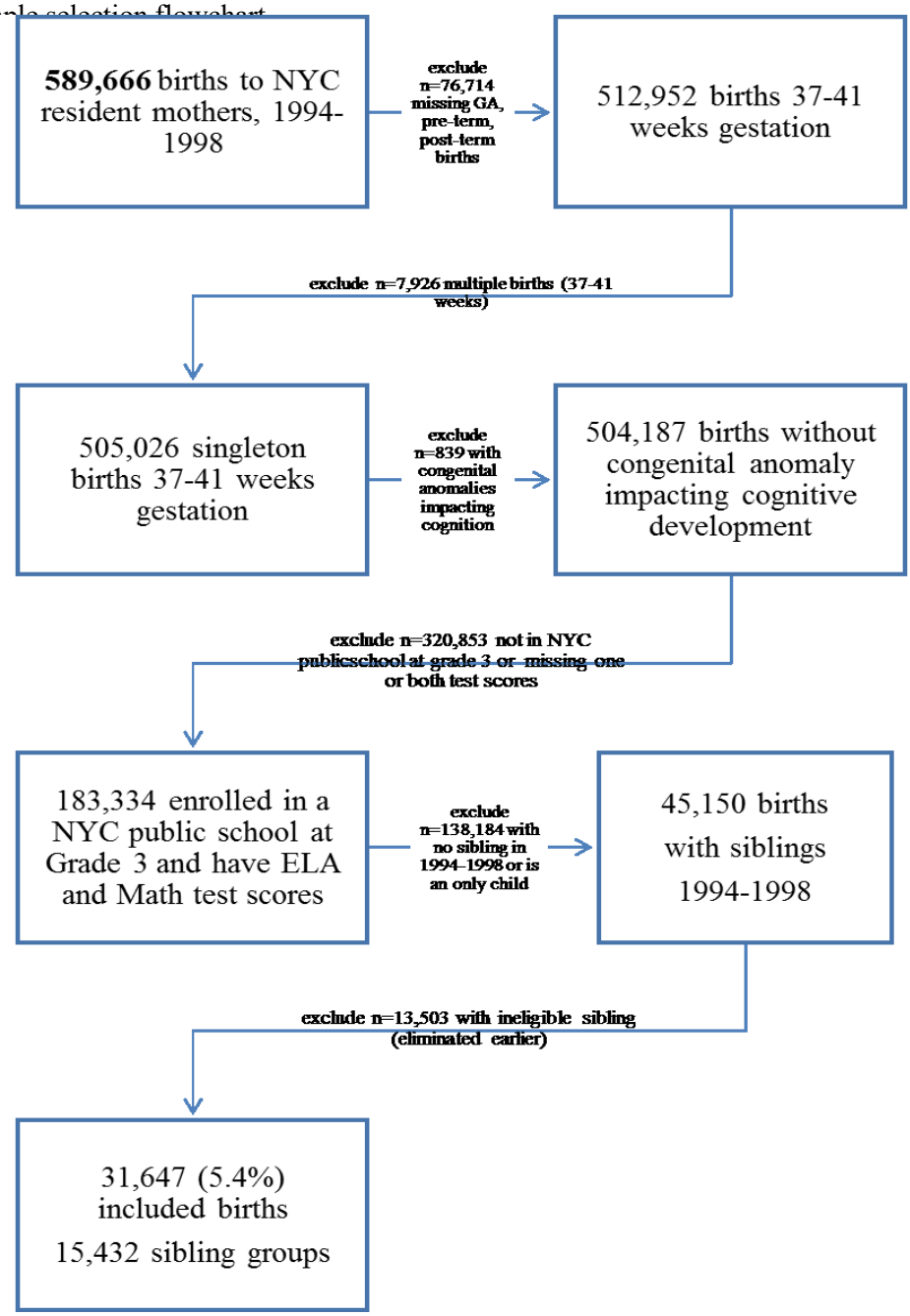


Table 4.3 presents demographic differences between included and excluded singleton births occurring at 37-41 weeks. Singleton births occurring at 37-41 weeks gestation who were included in the sample were more likely than those excluded to be non- white, and born earlier in the cohort (i.e., as birth year increased, the probability of being included decreased). Those included were also less likely to be non US born. Included and excluded births showed no difference with regards to mother's education completed, mother's age, mother's employment during pregnancy and neighborhood poverty level.

Table 4.3 Demographic Characteristics and Estimates* of being Included in Study Sample of Term (37-41 week gestation) Singleton Births to NYC Resident Mothers, 1994-1998 (n=505,026)

	Excluded (n=473,379)	Included (n=31,647)	GEE			
Parameter	**n (%)**	**n (%)**	**Estimate**	**Standard Error**	**95% CI**	**p value**
Intercept			1.563	0.57	0.44, 2.69	0.01
Mother's Race/Ethnicity						
black, non-Latino	132091 (27.90)	9700 (30.65)	0.011	0.00	0.008, 0.01	<.0001
Latino	164254 (34.70)	12116 (38.28)	0.014	0.00	0.01, 0.02	<.0001
Asian/Pacific Islander	47922 (10.12)	3633 (11.48)	0.018	0.00	0.016, 0.02	<.0001
Other	2259 (0.48)	266 (0.84)	0.042	0.01	0.03, 0.05	<.0001
white, non-Latino	125170 (26.44)	5932 (18.74)	*ref*			
Unknown/Missing	1683 (0.36)	0 (0.00)				
Mother's Education Completed						
Less than high school	117808 (24.89)	10778 (34.06)	*ref*			
High School	172190 (36.37)	11652 (36.82)	0.003	0.00	-0.0003, 0.01	0.07
Some College	84353 (17.82)	4992 (15.77)	0.001	0.00	-0.004, 0.01	0.84
4 years College or more	85091 (17.98)	3518 (11.12)	0.003	0.00	-0.002, 0.01	0.29
Missing	13937 (2.94)	707 (2.23)				
Mother's Age						
< 20 years old	49976 (10.56)	3664 (11.58)	*ref*			
20-34 years old	350802 (74.11)	24574 (77.65)	-0.006	0.00	-0.011, 0.00	0.03
35+ years old	72442 (15.30)	3407 (10.77)	-0.005	0.00	-0.012, 0.00	0.18
Missing	159 (0.03)	2 (0.01)				
Mother Employed During Pregnancy						
Yes	137686 (29.09)	7345 (23.21)	-0.001	0.00	-0.0023, 0.001	0.530
No	288152 (60.87)	21955 (69.37)	*ref*			
Missing	47541 (10.04)	2347 (7.42)				
Mother's Nativity						
US born	232923 (49.20)	16440 (51.95)	*ref*			
nonUS born	237729 (50.22)	15207 (48.05)	-0.008	0.00	-0.01, -0.01	<.0001
Missing	2727 (0.58)	0 (0.00)				
Census Tract poverty Level**						
<10%	118604 (25.05)	6495 (20.52)	*ref*			
10% to <20%	120160 (25.38)	7615 (24.06)	-0.002	0.00	-0.006, 0.00	0.44
20% to <30%	102328 (21.62)	7013 (22.16)	0.002	0.00	-0.004, 0.01	0.58
>= 30%	131357 27.75	10509 (33.21)	0.002	0.00	-0.004, 0.01	0.44
Missing	930 (0.20)	15 (0.05)				
BirthYear[+]			-0.001	0.00	-0.001, 0.00	0.01

*Estimates probability of being included in the study sample (vs excluded) using GEE (generalized estimating equation)
**% of residents living below 100% of the federal poverty level

[+]Child's year of birth included as a continuous variable

Demographic and pregnancy characteristics by gestational age are presented in Table 4.4. Since the odds of being born at 39-41 weeks gestation was computed with a random effects model (which accounts for the clustered nature of the data), a within sibling group interpretation of odds ratios can be made; however, since the non-partitioned version of the variables are being used, the estimates actually reflect a weighted average of the within and between sibling group estimates (if they are not the same) and therefore overestimate the true within sibling group value (80). Approximately 24% of births occurred early term. Compared to white, non-Latino births black non-Latino births were significantly less likely to occur at 39-41 weeks gestation (OR=0.82, 95%CI:0.75,0.90). Odds of birth at 39-41 weeks was non- significantly lower for Latino (OR=0.96, 95%CI: 0.89,1.05) and Asian/Pacific Islander (OR=0.93, 95%CI:0.83,1.05) and non-significantly higher for Other race/ethnicities (OR=1.26, 95%CI:0.89,1.79) compared to white non-Latino. There was no significant difference in the odds of birth at 39-41 weeks for mothers whose age was on average, 1 year above the mean age at birth (OR=0.99, 95%CI:0.99,1.00). Compared to births to mothers with a college education or higher, births to mothers with less than high school, high school, or some college had a significant and similar lower odds of giving birth at 39-41 weeks (OR= 0.87, 95%CI:0.79,0.97; OR=0.90, 95%CI:0.81,0.99);OR=0.88, 95%CI:0.79,0.99, respectively). Births to non-US born mothers were significantly more likely to occur at 39-weeks (OR=1.18, 95%CI: 1.11, 1.25). There was no significant difference in the odds of birth at 39-41 weeks for fathers whose age was on average, 1 year above the mean age at birth. (OR=1.00, 95%CI: 0.99, 1.00). There was no significant difference in the odds of birth at 39-41 weeks by father's level of education completed. Compared to a college education or higher, births to fathers with less than high school, high school, or some college was (OR= 0.94, 95%CI:0.84,1.05; OR=1.02, 95%CI:0.92,1.13); OR=1.02, 95%CI:0.90,1.15, respectively). There was no difference in the odds of birth at 39-41 weeks for children who subsequently were eligible for school lunch (a proxy measure of individual poverty) compared to those who were not (OR=0.93, 95% CI:0.85, 1.01). The odds of birth at 39-41 weeks was not significantly different

for mothers who were employed during pregnancy versus those who were not (OR=0.96, 95% CI:90,1.03). Financial coverage for pregnancy and birth was not associated with the odds of birth at 39-41 weeks. Compare to births covered by Medicaid, births covered by other forms of insurance or out-of-pocket payment did not have a significantly different odds of occurring at 39-41 weeks (HMO: OR=0.99, 95% CI:0.90, 1.10; Other 3rd Party: OR=0.95, 95% CI:0.88, 1.02; Self-Pay: OR= 1.03, 95% CI:0.90,1.19). Notably, there was a non-significantly higher odds of birth at 39-41 weeks for the overweight (OR=1.02, 95% CI: 0.90, 1.15) and obese (OR= 1.10, 95% CI:0.57, 2.13) weight categories compared to normal weight. Tobacco use, alcohol use and/or drug use all resulted in a lower odds of being born at 39-41 weeks, compared to when there was no substance use involved (OR=0.83, 95% CI:0.74,0.93; OR=0.67, 95% CI: 0.46,0.98; OR=0.63, 95% CI:0.49,0.81, respectively). Among clinical characteristics, having chronic (OR=0.44, 95% CI:0.25,0.78) or gestational diabetes (OR=0.67, 95% CI:0.58, 0.78) resulted in a lower odds of birth at 39-41 weeks compared to not having diabetes. Compared to not having pregnancy hypertension, having eclampsia (OR=0.33, 95% CI:0.09,1.26) or pregnancy-associated hypertension (OR=0.76, 95% CI:0.56, 1.04) resulted in lower, though non-significant odds of birth at 39-41 weeks; while having preeclampsia significantly decreased the odds of birth at 39-41 weeks (OR=0.54, 95% CI:0.44, 0.67). Chronic hypertension also decreased the odds of birth at 39-41 weeks (OR=0.49, 95% CI:0.35, 0.67). Among medical risk factors, Previous Preterm or SGA Infant (OR=0.46, 95% CI: 0.32, 0.66) and renal disease (OR=0.31, 95% CI:0.13,0.70) indicated a significantly lower odds of birth at 39-41 weeks; whereas other maternal risk factors suggested a higher odds of birth at 39-41 weeks (OR=1.11, 95% CI:1.02 , 1.21). With the exception of fetal distress which was associated with a higher odds of birth at 39-41 weeks (OR=1.20, 95% CI:1.01,1.43), all other complications of labor and delivery carried a significantly lower odds of birth at 39-41 weeks: cord prolapse (OR=0.94, 95% CI:0.44, 1.97); abruptio placenta (OR=0.44, 95% CI:0.27, 0.72); placenta previa (OR=0.29, 95% CI:0.18 , 0.47).

Delivery by C-section was associated with lower, though non-significant odds of birth at 39-41 weeks (OR=0.93, 95% CI: 0.86, 1.01). With the exception of births occurring in 1995(OR=0.93, 95% CI: 0.85, 1.02), births occurring in the remaining cohort years were significantly less likely to be 39-41 weeks gestation (1996: OR= 0.78, 95% CI:0.72, 0.86; 1997:OR=0.82, 95% CI:0.75,0.89; 1998:OR= 0.75, 95% CI:0.69,0.82). With respect to birth order, being born on average 1 place higher than the mean order was associated with a lower odds of birth at 39-41 weeks, i.e. children born later in the birth order were more likely to be born prior to 39 weeks (OR= 0.83, 95% CI:0.79, 0.87). Compared to females, children assigned male sex at birth had a lower odds of being born at 39-41 weeks gestation (OR=0.92, 95% CI:0.87, 0.97). For births with an Apgar score 1 unit higher than the mean, the odds of birth at 39-41 weeks were significantly higher (OR=1.07, 95% CI:1.02, 1.13). Births with any congenital anomaly (unrelated to cognitive development) had non-significant higher odds of birth at 39-41 weeks (OR=1.10, 95% CI: 0.70, 1.73). The odds of having been born at 39-41 weeks were non-significantly lower for children born small for gestational age (OR=0.97, 95% CI: 0.89, 1.06). NICU admission was associated with lower odds of having been born at 39-41 weeks (OR=0.69, 95% CI: 0.61, 0.78). There was no significant difference in the odds of having been born full term for those children who had a venous blood lead level above the mean (OR=1.00, 95% CI: 0.99,1.01). Finally, births with an ELA z-score higher than the average were more likely to have been born at 39-41 weeks (OR=1.05, 95% CI:1.02, 1.08). Similarly, births with a Math z-score higher than the average were more likely to have been born at 39-41 weeks (OR=1.05, 95% CI:1.02, 1.08).

Table 4.4 Demographic and Clinical Characteristics of Study Sample Births, by Gestational Age Group (n=31,647)

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
	n	%	n	%	n	%			
	31,647	100	7733	24.44	23914	75.56			
Mother's Race/Ethnicity									
black, non-Latino	9700	30.65	2560	33.10	7140	29.86	0.82	0.75	0.90
Latino	12116	38.28	2868	37.09	9248	38.67	0.96	0.89	1.05
Asian/Pacific Islander	3633	11.48	883	11.42	2750	11.50	0.93	0.83	1.04
Other	266	0.84	52	0.67	214	0.89	1.26	0.89	1.79
white, non-Latino	5932	18.74	1370	17.72	4562	19.08	*ref*		
Mother's Age[b]									
mean (sd)	26.83	(5.87)	27.02	(5.97)	26.77	(5.83)	0.99	0.99	1.00
Mother's Education Completed									
Less than high school	10778	34.06	2687	34.75	8091	33.83	0.87	0.79	0.97
High School	11652	36.82	2834	36.65	8818	36.87	0.90	0.81	0.99
Some College	4992	15.77	1229	15.89	3763	15.74	0.88	0.79	0.99
4 years College or more	3518	11.12	796	10.29	2722	11.38	*ref*		
Missing	707	2.23	187	2.42	520	2.17			
Mother's Nativity									
US born	16440	51.95	4235	54.77	12205	51.04	*ref*		
nonUS born	15207	48.05	3498	45.23	11709	48.96	1.18	1.11	1.25
Father's Age[b]									
mean (sd)	31.06	(6.86)	31.13	(6.93)	31.03	(6.84)	1.00	0.99	1.00

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
Father's Education Completed									
Less than high school	5640	17.82	1413	18.27	4227	17.68	0.94	0.84	1.05
High School	10302	32.55	2448	31.66	7854	32.84	1.02	0.92	1.13
Some College	3937	12.44	936	12.1	3001	12.55	1.02	0.90	1.15
4 years College or more	3480	11	838	10.84	2642	11.05	*ref*		
Missing	8288	26.19	2098	27.13	6190	25.88			
Was child ever eligible for school lunch program**									
Yes	27181	85.89	6698	86.62	20483	85.65	0.93	0.85	1.01
No	4466	14.11	1035	13.38	3431	14.35	*ref*		
Mother Employed During This Pregnancy									
Yes	7345	23.21	1836	23.74	5509	23.04	0.96	0.90	1.03
No	21955	69.37	5347	69.15	16608	69.45	*ref*		
Missing	2347	7.42	550	7.11	1797	7.51			
Primary Financial Coverage for Pregnancy and Birth									
HMO	3019	9.54	733	9.48	2286	9.56	0.99	0.90	1.10
Other 3rd Party	7155	22.61	1797	23.24	5358	22.41	0.95	0.88	1.02
Self-pay	1369	4.33	323	4.18	1046	4.37	1.03	0.90	1.19
Medicaid	19487	61.58	4718	61.01	14769	61.76	*ref*		
Missing	617	1.95	162	2.09	455	1.9			
Mother's Weight***									
Normal weight	24784	78.31	6018	77.82	18766	78.47	*ref*		
Overweight	1943	6.14	467	6.04	1476	6.17	1.02	0.90	1.15
Obese	65	0.21	14	0.18	51	0.21	1.10	0.57	2.13

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
Missing	4855	15.34	1234	15.96	3621	15.14			
Mother's tobacco use during pregnancy									
Tobacco use	1970	6.22	545	7.05	1425	5.96	0.83	0.74	0.93
No tobacco use	29289	92.55	7088	91.66	22201	92.84	*ref*		
Missing	388	1.23	100	1.29	288	1.2			
Mother's Alcohol Use During Pregnancy									
Yes	155	0.49	50	0.65	105	0.44	0.67	0.46	0.98
No	30853	97.49	7520	97.25	23333	97.57	*ref*		
Missing	639	2.02	163	2.11	476	1.99			
Mother's Use of Amphetamines, cocaine, heroin, methadone or marijuana									
Yes	373	1.18	126	1.63	247	1.03	0.63	0.49	0.81
No	31274	98.82	7607	98.37	23667	98.97	*ref*		
Diabetes									
Chronic diabetes	62	0.2	26	0.34	36	0.15	0.44	0.25	0.78
Gestational diabetes	1090	3.44	343	4.44	747	3.12	0.67	0.58	0.78
No diabetes	30495	96.36	7364	95.23	23131	96.73	*ref*		
Pregnancy-related hypertension									
Eclampsia	11	0.03	5	0.06	6	0.03	0.33	0.09	1.26
Preeclampsia	459	1.45	164	2.12	295	1.23	0.54	0.44	0.67
Pregnancy-associated hypertension	251	0.79	72	0.93	179	0.75	0.76	0.56	1.04

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
No pregnancy-related hypertension	30926	97.72	7492	96.88	23434	97.99	*ref*		
Chronic hypertension									
Yes	211	0.67	82	1.06	129	0.54	0.49	0.35	0.67
No	31436	99.33	7651	98.94	23785	99.46	*ref*		
No prenatal care received									
Yes	740	2.34	198	2.56	542	2.27	0.86	0.72	1.03
No	28569	90.27	6856	88.66	21713	90.8	*ref*		
Missing	2338	7.39	679	8.78	1659	6.94			
Medical Risk Factor - Previous Preterm or SGA Infant									
Yes	160	0.51	65	0.84	95	0.4	0.46	0.32	0.66
No	31338	99.02	7634	98.72	23704	99.12	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Medical Risk Factor - Renal Disease									
Yes	31	0.1	15	0.19	16	0.07	0.31	0.13	0.70
No	31467	99.43	7684	99.37	23783	99.45	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Other maternal risk factors									
At least 1 misc. maternal risk factor	4185	13.22	964	12.47	3221	13.47	1.11	1.02	1.21
None of the misc. Maternal risk factors	27462	86.78	6769	87.53	20693	86.53	*ref*		
Complication of Labor/Delivery - Cord Prolapse									
Yes	46	0.15	12	0.16	34	0.14	0.94	0.44	1.97

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
No	31452	99.38	7687	99.41	23765	99.38	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Complication of Labor/Delivery - Fetal Distress									
Yes	920	2.91	199	2.57	721	3.01	1.20	1.01	1.43
No	30578	96.62	7500	96.99	23078	96.5	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Complication of Labor/Delivery - Abruption Placenta									
Yes	86	0.27	35	0.45	51	0.21	0.44	0.27	0.72
No	31412	99.26	7664	99.11	23748	99.31	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Complication of Labor/Delivery - Placenta Previa									
Yes	85	0.27	42	0.54	43	0.18	0.29	0.18	0.47
No	31413	99.26	7657	99.02	23756	99.34	*ref*		
Missing	149	0.47	34	0.44	115	0.48			
Method of Delivery									
Vaginal	26174	82.71	6364	82.3	19810	82.84	*ref*		
C-section	4987	15.76	1275	16.49	3712	15.52	0.93	0.86	1.01
Unknown/Missing	486	1.54	94	1.22	392	1.64			
Child's year of birth									
1994	6337	20.02	1381	17.86	4956	20.72	*ref*		
1995	5929	18.73	1353	17.5	4576	19.14	0.93	0.85	1.02
1996	6559	20.73	1696	21.93	4863	20.34	0.78	0.72	0.86

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
1997	6070	19.18	1519	19.64	4551	19.03	0.82	0.75	0.89
1998	6752	21.34	1784	23.07	4968	20.77	0.75	0.69	0.82
Birth Order[b]									
1st	15102	47.72	3365	43.51	11737	49.08			
2nd	15116	47.76	4002	51.75	11114	46.47			
3rd or higher	1429	4.51	366	4.72	1063	4.44			
mean(sd)	1.57	(0.59)	1.62	(0.59)	1.56	(0.59)	0.83	0.79	0.87
Child's Sex assigned a birth									
Female	16327	51.59	3883	50.21	12444	52.04	*ref*		
Male	15320	48.41	3850	49.79	11470	47.96	0.92	0.87	0.97
Apgar Score, 5 Minutes[b]									
<7	142	0.45	43	0.56	99	0.41			
7+	31308	98.93	7637	98.76	23671	98.98			
Unknown/Missing	197	0.62	53	0.69	144	0.6			
mean (sd)	9.11	(0.58)	9.1	(0.62)	9.1	(0.57)	1.07	1.02	1.13
Any Congenital Anomaly									
Yes	131	0.41	30	0.39	101	0.42	1.10	0.70	1.73
No	31516	99.59	7703	99.61	23813	99.58	*ref*		
Small for Gestational Age (<10th percentile of birthweight for gestational age)									
Yes	4095	12.94	1045	13.51	3050	12.75	0.97	0.89	1.06
No	27552	87.06	6688	86.49	20864	87.25	*ref*		

Characteristic	Total Sample		37-38 weeks		39-41 weeks		Odds Ratio[a]	95% CI	
Neonatal Intensive Care Unit Admission									
Yes	1558	4.92	493	6.38	1065	4.45	0.69	0.61	0.78
No	24018	75.89	5933	76.72	18085	75.63	*ref*		
Missing	6071	19.18	1307	16.9	4764	19.92			
Highest venous blood-lead level test[b]									
mean (sd)	4.87	(3.84)	4.9	(3.87)	4.9	(3.83)	1.00	0.99	1.01
Child's earliest 3rd grade English Language Arts (ELA) z-score[b]									
mean (sd)	0	(1.00)	-0.04	(1.02)	0.014	(1.00)	1.05	1.02	1.08
Child's earliest 3rd grade Math z-score[b]									
mean (sd)	0.07	(0.98)	0.04	(0.99)	0.08	(0.98)	1.05	1.02	1.08

**a measure of individual level poverty

***Only weight information was available on the medical report accompanying birth certificate submissions; therefore, Body Mass Index (BMI) could not be calculated.

[a]Odds of being born at gestational age 39-41 weeks

[b]Effects of continuous variables are assessed as one unit offsets from the mean.

The relationship between covariates and ELA and math z-scores are presented in Table 4.5. Compared to white, non-Latino births black non-Latino births, Latino births and Other race/ethnicities had significantly lower ELA and Math z-score by 0.80 units, 0.70 units and 0.26 units for ELA and 0.73 units, 0.55 units and 0.14 units for Math, respectively. Asian/Pacific Islander had significantly higher ELA and Math z-scores by 0.05 units compared to white non-Latino. In general, an increase in mother's age (ELA-β=0.03, 95% CI:0.03,0.03; Math- β=0.03, 95% CI:0.02,0.03) or father's age (ELA-β=0.02, 95% CI:0.01,0.02; Math- β=0.02, 95% CI:0.01,0.02), being born to non US born woman (ELA-β=0.21, 95% CI:0.18,0.24; Math-β=0.32, 95% CI:0.29,0.34), employment during pregnancy (ELA-β=0.37, 95% CI:0.34,0.39; Math- β=0.31, 95% CI:0.28,0.33), and non-Medicaid payment for birth were associated with significantly higher ELA and Math z-scores. Conversely, having less than a college degree, child poverty (as measured by eligibility for school lunch) (ELA-β=-0.56, 95% CI:-0.59,-0.52; Math-β=-0.44, 95% CI:-0.48,-0.41), births among those with overweight (ELA-β=-0.19, 95% CI:-0.24,-0.14; Math- β=-0.23, 95% CI:-0.28,-0.18) or obese weight (ELA-β=-0.57, 95% CI:-0.82,-0.32; Math- β=-0.50, 95% CI:-0.74,-0.25), tobacco (ELA-β=-0.27, 95% CI:-0.32,-0.23; Math- β=-0.28, 95% CI:-0.33,-0.23), alcohol (ELA-β=-0.31, 95% CI:-0.45,-0.16; Math- β=-0.31, 95% CI:-0.46,-0.17) or other substance use (ELA-β=-0.32, 95% CI:-0.42,-0.22; Math- β=-0.30, 95% CI:-0.40,-0.20) were associated with significantly lower ELA and Math z-scores. Among clinical characteristics, no prenatal care received (ELA-β=-0.18, 95% CI:-0.25,-0.11; Math- β=-0.17, 95% CI:-0.24,-0.10), chronic hypertension (ELA-β=-0.22, 95% CI:-0.36,-0.09; Math- β=-0.16, 95% CI:-0.29,-0.03), having had a previous preterm or small for gestational age infant (ELA-β=-0.27, 95% CI:-0.42,-0.12; Math- β=-0.28, 95% CI:-0.43,-0.13), if the current birth was small for gestational age (ELA-β=-0.10, 95% CI:-0.13,-0.07; Math- β=-0.11, 95% CI:-0.14,-0.08), or if there was an indication for admission to NICU (ELA-β=-0.17, 95% CI:-0.21,-0.12; Math- β=-0.18, 95% CI:-0.43,-0.13) was also associated with significantly lower ELA and Math z-scores. Having a higher Apgar score (ELA-β= 0.05, 95% CI: 0.03, 0.06; Math- β=0.05, 95% CI: 0.03,

0.07) and being delivered by C-section (ELA-β= 0.08, 95% CI: 0.05, 0.11; Math- β=0.06, 95% CI: 0.03, 0.09) were both associated with significantly higher ELA and Math z-scores. Males were significantly more likely to have lower ELA z-scores compared to females (ELA-β=-0.18, 95% CI:-0.20,-0.16), but they did not differ on Math (β=0.00, 95% CI:-0.02,0.02) Lastly, being born later in the birth order (ELA-β= -0.12, 95% CI:-0.13,-0.10; Math- β=-0.08, 95% CI:-0.10,-0.07) and increases in venous blood lead level (ELA-β=-0.03, 95% CI:-0.03,-0.02; Math- β=-0.02, 95% CI:-0.02,-0.02) were associated with a significant decrease in Math and ELA z-scores.

Table 4.5 English Language Arts (ELA) and Math z-scores by Demographic and Clinical Characteristics of Study Sample Births (n=31,647)

	ELA			Math		
Characteristic	**β**	**95% CI**		**β**	**95% CI**	
Mother's Race/Ethnicity						
black, non-Latino	-0.80	-0.84	-0.77	-0.73	-0.76	-0.69
Latino	-0.70	-0.73	-0.66	-0.55	-0.59	-0.52
Asian/Pacific Islander	0.05	0.01	0.10	0.28	0.23	0.32
Other	-0.26	-0.39	-0.12	-0.14	-0.27	-0.01
white, non-Latino	*ref*			*ref*		
Mother's Age	0.03	0.03	0.03	0.03	0.02	0.03
Mother's Education Completed						
Less than high school	-0.89	-0.93	-0.85	-0.72	-0.76	-0.68
High School	-0.62	-0.66	-0.58	-0.49	-0.53	-0.45
Some College	-0.46	-0.50	-0.41	-0.38	-0.42	-0.33
4 years College or more	*ref*			*ref*		
Mother's Nativity						
US born	*ref*			*ref*		
nonUS born	0.21	0.18	0.24	0.32	0.29	0.34
Father's Age	0.02	0.01	0.02	0.02	0.01	0.02
Father's Education Completed						
Less than high school	-0.75	-0.79	-0.71	-0.62	-0.66	-0.57
High School	-0.50	-0.54	-0.46	-0.42	-0.46	-0.38
Some College	-0.35	-0.39	-0.30	-0.30	-0.35	-0.26
4 years College or more	*ref*			*ref*		

Characteristic	ELA β	ELA 95% CI		Math β	Math 95% CI	
Was child ever eligible for school lunch program**						
Yes	-0.56	-0.59	-0.52	-0.44	-0.48	-0.41
No	*ref*			*ref*		
Mother Employed During This Pregnancy						
Yes	0.37	0.34	0.39	0.31	0.28	0.33
No	*ref*			*ref*		
Primary Financial Coverage for Pregnancy and Birth						
HMO	0.37	0.33	0.40	0.29	0.25	0.32
Other 3rd Party	0.50	0.47	0.53	0.42	0.40	0.45
Self-pay	0.19	0.13	0.24	0.19	0.14	0.24
Medicaid	*ref*			*ref*		
Mother's Weight***						
Normal weight	*ref*			*ref*		
Overweight	-0.19	-0.24	-0.14	-0.23	-0.28	-0.18
Obese	-0.57	-0.82	-0.32	-0.50	-0.74	-0.25
Mother's tobacco use during pregnancy						
Tobacco use	-0.27	-0.32	-0.23	-0.28	-0.33	-0.23
No tobacco use	*ref*			*ref*		
Mother's Alcohol Use During Pregnancy						
Yes	-0.31	-0.45	-0.16	-0.31	-0.46	-0.17
No	*ref*			*ref*		

Characteristic	ELA β	ELA 95% CI		Math β	Math 95% CI	
Mother's Use of Amphetamines, cocaine, heroin, methadone or marijuana						
Yes	-0.32	-0.42	-0.22	-0.30	-0.40	-0.20
No	*ref*			*ref*		
Diabetes						
Chronic diabetes	-0.14	-0.38	0.10	0.00	-0.23	0.23
Gestational diabetes	0.002	-0.06	0.06	-0.01	-0.07	0.05
Pregnancy-related hypertension						
Eclampsia	-0.56	-1.10	-0.01	-0.18	-0.71	0.36
Preeclampsia	-0.11	-0.20	-0.03	-0.08	-0.17	0.00
Pregnancy-associated hypertension	-0.07	-0.19	0.04	-0.09	-0.21	0.02
No pregnancy-related hypertension	*ref*			*ref*		
Chronic hypertension						
Yes	-0.22	-0.36	-0.09	-0.16	-0.29	-0.03
No	*ref*			*ref*		
No prenatal care received						
Yes	-0.18	-0.25	-0.11	-0.17	-0.24	-0.10
No	*ref*			*ref*		
Medical Risk Factor - Previous Preterm or SGA Infant						
Yes	-0.27	-0.42	-0.12	-0.28	-0.43	-0.13
No	*ref*			*ref*		
Medical Risk Factor - Renal Disease						

	ELA			Math		
Characteristic	**β**	**95% CI**		**β**	**95% CI**	
Yes	0.01	-0.33	0.35	0.08	-0.26	0.41
No	*ref*			*ref*		
Other maternal risk factors						
At least 1 misc. maternal risk factor	-0.02	-0.05	0.01	-0.03	-0.06	0.00
None of the misc. Maternal risk factors	*ref*			*ref*		
Complication of Labor/Delivery - Cord Prolapse						
Yes	-0.21	-0.48	0.06	-0.25	0.52	0.02
No	*ref*			*ref*		
Complication of Labor/Delivery - Fetal Distress						
Yes	-0.03	-0.09	0.03	-0.01	-0.07	0.05
No	*ref*			*ref*		
Complication of Labor/Delivery - Abruption Placenta						
Yes	-0.11	-0.30	0.09	-0.28	-0.47	-0.08
No	*ref*			*ref*		
Complication of Labor/Delivery - Placenta Previa						
Yes	-0.08	-0.28	0.12	-0.07	-0.27	0.12
No	*ref*			*ref*		
Method of Delivery						
Vaginal	*ref*			*ref*		
C-section	0.08	0.05	0.11	0.06	0.03	0.09
Unknown/Missing						

	ELA			Math		
Characteristic	**β**	**95% CI**		**β**	**95% CI**	
Child's year of birth						
1994	*ref*			*ref*		
1995	-0.04	-0.07	-0.01	-0.02	-0.05	0.01
1996	-0.07	-0.10	-0.04	-0.05	-0.08	-0.02
1997	-0.09	-0.12	-0.06	-0.01	-0.04	0.02
1998	-0.11	-0.14	-0.08	-0.09	-0.12	-0.06
Birth Order	-0.12	-0.13	-0.10	-0.08	-0.10	-0.07
Child's Sex assigned a birth						
Female	*ref*			*ref*		
Male	-0.18	-0.20	-0.16	0.00	-0.02	0.02
Apgar Score, 5 Minutes	0.05	0.03	0.06	0.05	0.03	0.07
Any Congenital Anomaly						
Yes	-0.21	-0.37	-0.05	-0.12	-0.28	0.04
No	*ref*			*ref*		
Small for Gestational Age (<10th percentile of birthweight for gestational age)						
Yes	-0.10	-0.13	-0.07	-0.11	-0.14	-0.08
No	*ref*			*ref*		
Neonatal Intensive Care Unit Admission						
Yes	-0.17	-0.21	-0.12	-0.18	-0.23	-0.13
No	*ref*			*ref*		
Highest venous blood-lead level test	-0.03	-0.03	-0.02	-0.02	-0.02	-0.02

a measure of individual level poverty;*Only weight information was available on the medical report accompanying birth certificate submissions; therefore, Body Mass Index (BMI) could not be calculated. [a]Odds of being born at gestational age 39-41 weeks. [b]Effects of continuous variables are assessed as one unit offsets from the mean.

Multiple Imputation

Multiple imputation was performed on covariates with missing data. The frequency of missingness for each variable and the overall patterns are presented in tables 4.6 and 4.7, respectively. Father's education had the largest proportion of missing data (26.19%). The 25 multiply imputed data sets created resulted in 791,175 observations (31,647*25). The group means for each variable show some variation according to the pattern of missingness (i.e. one can predict missingness of a variable from other variables in the dataset), suggesting that the assumption of a mechanism other than MCAR is reliable. Overall, 590 data patterns were observed with a range of missing data on 1 to 12 covariates. Only 1 record was missing data on 12 covariates. A total of 36.78% of records had complete information for all of the covariates included in this analysis. Trace plots were generated for the following imputed continuous variables (5-minute Apgar score, highest venous blood lead level test result, and father's age at last birthday). The trace plots, which can be found in Appendix B, show random patterns which indicate good quality of the imputations. (104)

Appendix B also includes a comparison of means, and standard errors of imputed variables versus values from the original data. Results indicate a strong consistency and no major aberrations in values. Additional frequencies by imputation show that for each imputation, the values of the imputed variables show strong consistency in distribution with that seen in the original data set. Notably, for father's education, there was an overestimate of the percent '< high school' in the imputed data sets versus the original (27.33% versus 17.82%, respectively), and slight overestimate of 'some college' and 'college or more' versus the original (some college, 14.65% versus 12.44%; college or more, 13.45% versus 11%, respectively). This is likely due to the large number of missing that existed for this variable.

Table 4.6 Frequency of Missing Values for Covariates among Births Included in the Study Sample (n=31.647)

Variable	N	N missing	% Missing
Father's Education Completed	23359	8288	26.19%
Father's Age	24034	7613	24.06%
Highest venous blood lead level test	24785	6862	21.68%
Neonatal Intensive Care Unit Admission	25576	6071	19.18%
Mother's Weight	26792	4855	15.34%
Mother Employed During This Pregnancy	29300	2347	7.42%
No prenatal care received	30000	1647	5.20%
Mother's Education Completed	30940	707	2.23%
Mother's Alcohol Use During Pregnancy	31008	639	2.02%
Primary Financial Coverage for Pregnancy and Birth	31030	617	1.95%
Method of Delivery	31161	486	1.54%
Mother's tobacco use during pregnancy	31259	388	1.23%
Apgar Score,5 Minutes	31450	197	0.62%
Percent Persons African American in census tract*	31488	159	0.50%
Medical Risk Factor - Previous Preterm or SGA infant	31498	149	0.47%
Medical Risk Factor - Renal Disease	31498	149	0.47%
Complication of Labor/Delivery - Cord Prolapse	31498	149	0.47%
Complication of Labor/Delivery Fetal Distress	31498	149	0.47%
Complication of Labor/Delivery Abruption Placenta	31498	149	0.47%
Complication of Labor/Delivery Placenta Previa	31498	149	0.47%
Census Tract poverty level*	31632	15	0.05%
Mother's Age	31647	0	0.00%
Diabetes	31647	0	0.00%
Pregnancy-related hypertension	31647	0	0.00%
Other maternal risk factors	31647	0	0.00%
Chronic hypertension	31647	0	0.00%
Mother's Use of Amphetamines, cocaine, heroin, methadone or marijuana	31647	0	0.00%
Small for Gestational Age (<10th percentile of birthweight for gestational age)	31647	0	0.00%
Congenital Anomaly	31647	0	0.00%
Gestational Age at Delivery	31647	0	0.00%
Mother's Race/Ethnicity	31647	0	0.00%
Child's Sex assigned a birth	31647	0	0.00%
Birth Order	31647	0	0.00%
Mother's Nativity	31647	0	0.00%
Child's year of birth	31647	0	0.00%
Was child ever eligible for school lunch program	31647	0	0.00%
Earliest 3rd grade ELA test score (z-score)	31647	0	0.00%
Earliest 3rd grade Math test score (z-score)	31647	0	0.00%

*Auxiliary variables

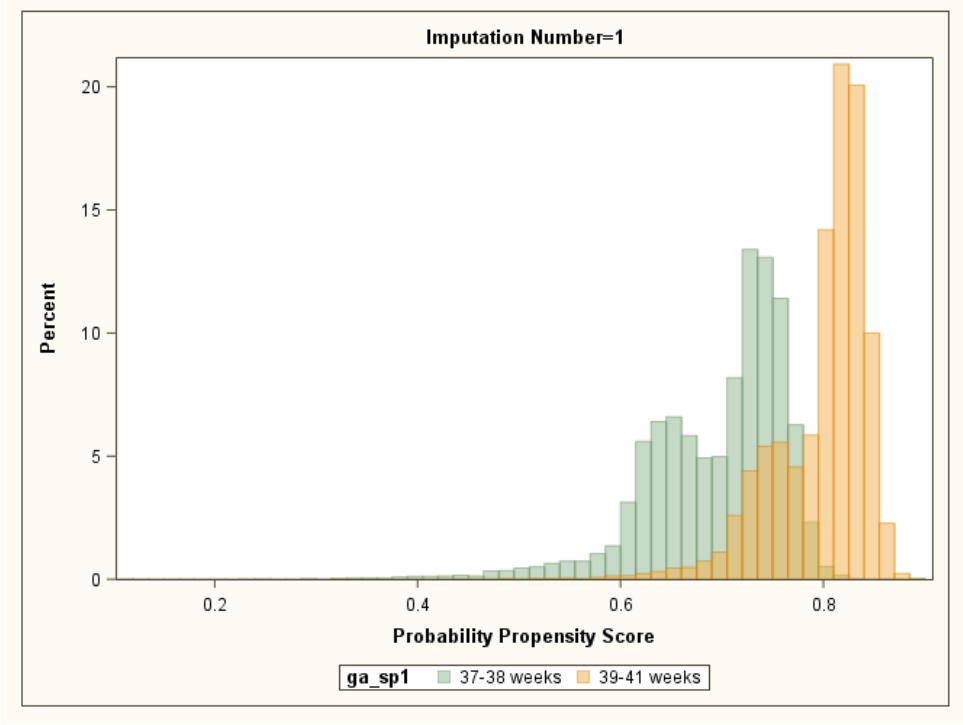
Table 4.7 Count of Missingness Patterns

# variables with missing data	# records	Percent
0	11640	36.78
1	7612	24.05
2	6499	20.54
3	3468	10.96
4 or more	2428	7.67
	31647	100.00

Propensity Scores

Figure 4.4 displays a plot of the overlap in predicted probabilities as well as the overlap in log odds of birth at 39-41 weeks between early term (37-38) and full term (39-41) births for imputation 1. Tables 4.8 and 4.9 show the distribution of ranks of predicted probabilities and ranks of log odds propensity scores for early term (37-38) and full term (39-41) births. Among the predicted probabilities, while there is a modest region of overlap, there are substantial numbers of non-overlapping observations, suggesting that strong heterogeneity in the distribution of covariates between early term and full term births remains. This is confirmed by the variation in the distribution of early term and full term births within the ranks (based on quintiles) of the predicted probabilities. As the propensity score is a 'balancing score' (97), covariate adjustment using the predicted probabilities may not be of appropriate use for this data set. The log odds propensity score provided greater balance between the early term and full term categories, and will be used in test score models. Use of the log odds propensity score for propensity score stratification has occurred in educational settings, as has the use of stratified ranks for covariate adjustment. (95,98,105) Appendix C provides additional summaries displaying the overlap in log odds propensity between these groups for the other imputed data sets.

Figure 4.4 Overlap in Predicted Probabilities and Log Odds Propensity Score for Early Term (37-38 weeks) and Full Term (39-41 weeks) Births



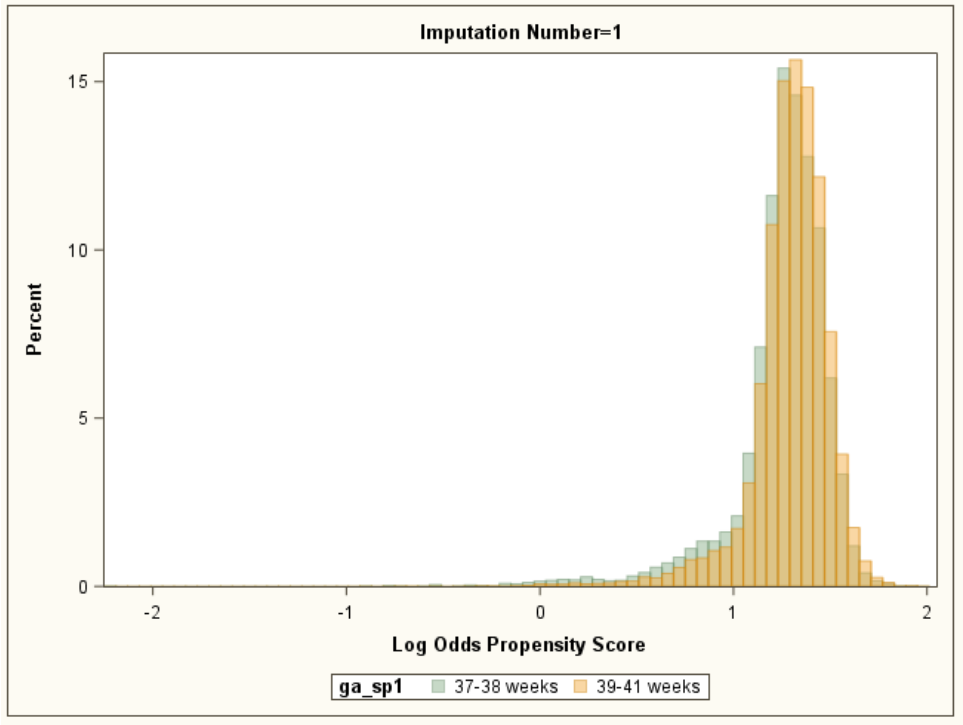


Table 4.8. Distribution of Early Term (37-38 weeks) and Full Term (39-41 weeks) Births by Predicted Probability Quintile Ranks

Predicted Probability Quintile Ranks	**Gestational Age n (%)**		
	37-38 weeks	**39-41 weeks**	**Total**
0	4418 (57.13)	1911 (7.99)	6329
1	2756 (35.64)	3574 (14.95)	6330
2	543 (7.02)	5786 (24.20)	6329
3	14 (0.18)	6316 (26.41)	6330
4	2 (0.03)	6327 (26.46)	6329
Total	7733 (24.44)	23914 (75.56)	31647 100

Table 4.9. Distribution of Early Term (37-38 weeks) and Full Term (39-41 weeks) Births by Log Odds Quintile Ranks

Log Odds Quintile Ranks	**Gestational Age n (%)**		
	37-38 weeks	**39-41 weeks**	**Total**
0	1928 (24.93)	4401 (18.40)	6329
1	1602 (20.72)	4728 (19.77)	6330
2	1497 (19.36)	4832 (20.21)	6329
3	1405 (18.17)	4925 (20.59)	6330
4	1301 (16.82)	5028 (21.03)	6329
Total	7733 (24.44)	23914 (75.56)	31647 100

Summary estimates calculated across imputed datasets using Proc MIAnalyze are shown in Table 4.10. The Level 2 error variance estimate from the null model (Model 0, containing no covariates) was identical across imputations [0.5596]; this is expected since there are no covariates in the model. This estimate was used to calculate an intraclass correlation coefficient (ICC) of 0.1454 [0.5596/ (0.5596+3.29)]. Thus, nearly 15% of the variability in gestational age is accounted for by the sibling groups in the study sample. The results also show that the variability in the log odds of birth at 39-41 weeks between sibling groups in the study sample is statistically significant [p<.0001]. This confirms that the rate of birth at 39-41 weeks varies across sibling groups. The fixed effects intercept estimate computed from the null model was 1.2734. This estimate is the log odds of birth at 39-41 weeks for the average sibling group in the sample when all other variables =0. Using this estimate, the probability of birth at 39-41 weeks for an average sibling group in the study sample is $0.7813(e^{1.2734}/1+e^{1.2734})$.

In the model containing demographic and clinical variables used to predict gestational age 'assignment' (37-38 or 39-41 weeks) (Model 1, in Table 4.10), within sibling group, an individual sibling's deviation from the sibling group average for resulted in significantly lower probability of birth at full term for mother's age (β=-0.061, 95%CI:-0.089,-0.032), preeclmapsia (β=-0.205,95%CI:-0.315,-0.095) and placenta previa (β=-1.198,95%CI:-1.897,-0.500). In other words, for a given sibling, a within sibling group increase in mother's age, having preeclmapsia, or placenta previa is significantly associated with lower odds of full term birth. Between sibling groups, a change in the average employment status is significantly associated with a decrease in the probability of being born full term. Tobacco use (β=-0.163,95%CI:-0.317,-0.009), preeclampsia (β=-0.148,95%CI:-0.258,-0.039), chronic hypertension (β=-0.966,95%CI:-1.404,-0.528) , having had a previous preterm or small for gestational age birth (β=-1.230,95%CI:-1.739,-0.722), renal disease (β=-1.333,95%CI:-2.537,-0.129), drug use during pregnancy

(β=-0.629,95%CI:-0.317,-0.009), abruptio placenta (β=-0.842,95%CI:-1.608,-0.075), and placenta previa (β= -1.210, 95%CI:-1.976,-0.445) were all associated with a lower odds of birth at full term. Notably, a change in the average education level (β=0.084,95%CI:0.037,0.132) for a sibling group is significantly associated with an increased probability of full term birth. In other words, higher education is associated with being born at full term.

Table 4.10 Multilevel Predictors of Birth at 39-41 Weeks of Gestational Age among Study Sample Births, Over 25 Imputations.

		Model 0			Model 1				
	Parameter	**Estimate**	**Standard Error**	**95% CI**	**Estimate**	**Standard Error**	**95% CI**		
	Intercept	1.273	0.020		1.290	0.084	1.126	1.454	***
Mother's Age	**Within**				-0.061	0.015	-0.089	-0.032	***
	Between				-0.008	0.004	-0.016	0.000	
Father's Age	**Within**				-0.008	0.008	-0.024	0.009	
	Between				0.007	0.004	0.000	0.014	
Mother Employed During This Pregnancy	**Within**				-0.045	0.066	-0.174	0.085	
	Between				-0.105	0.052	-0.206	-0.004	*
Primary Financial Coverage for Pregnancy and Birth	**Within**				-0.009	0.029	-0.066	0.047	
	Between				-0.035	0.022	-0.078	0.009	
Mother's tobacco use during pregnancy	**Within**				0.045	0.101	-0.153	0.244	
	Between				-0.163	0.079	-0.317	-0.009	*
Diabetes	**Within**				-0.094	0.062	-0.216	0.027	
	Between				-0.260	0.049	-0.356	-0.164	***
Pregnancy-related hypertension	**Within**				-0.205	0.056	-0.315	-0.095	***
	Between				-0.148	0.056	-0.258	-0.039	**
Other maternal risk factors	**Within**				0.026	0.062	-0.096	0.147	
	Between				0.100	0.067	-0.031	0.230	
Chronic hypertension	**Within**				-0.192	0.250	-0.682	0.298	
	Between				-0.966	0.223	-1.404	-0.528	***
Medical Risk Factor - Previous Preterm or SGA Infant	**Within**				-0.059	0.263	-0.575	0.457	
	Between				-1.230	0.259	-1.739	-0.722	***
Medical Risk Factor - Renal Disease	**Within**				-1.067	0.595	-2.233	0.100	
	Between				-1.333	0.614	-2.537	-0.129	*
Mother's Use of	**Within**				0.278	0.213	-0.139	0.696	

		Model 0			Model 1				
	Parameter	**Estimate**	**Standard Error**	**95% CI**	**Estimate**	**Standard Error**	**95% CI**		
Amphetamines, cocaine, heroin, methadone or marijuana	**Between**				-0.629	0.191	-1.004	-0.255	**
Complication of Labor/Delivery - Cord Prolapse	**Within**				0.147	0.494	-0.822	1.116	
	Between				-0.342	0.546	-1.413	0.729	
Complication of Labor/Delivery - Fetal Distress	**Within**				0.215	0.120	-0.020	0.450	
	Between				0.222	0.128	-0.028	0.473	
Complication of Labor/Delivery - Abruption Placenta	**Within**				-0.627	0.343	-1.299	0.046	
	Between				-0.842	0.391	-1.608	-0.075	*
Complication of Labor/Delivery - Placenta Previa	**Within**				-1.198	0.356	-1.897	-0.500	***
	Between				-1.210	0.390	-1.976	-0.445	**
No prenatal care received	**Within**				-0.056	0.132	-0.315	0.203	
	Between				-0.188	0.132	-0.447	0.072	
Mother's Alcohol Use During Pregnancy	**Within**				-0.141	0.263	-0.657	0.374	
	Between				-0.129	0.276	-0.670	0.413	
Mother's Education Completed	**Within**				-0.047	0.044	-0.133	0.039	
	Between				0.084	0.024	0.037	0.132	***
Father's Education Completed	**Within**				0.025	0.042	-0.057	0.107	
	Between				-0.005	0.026	-0.055	0.045	
Any Congenital Anomaly	**Within**				0.060	0.312	-0.552	0.671	
	Between				0.190	0.336	-0.469	0.849	
Mother's Weight	**Within**				0.012	0.116	-0.215	0.240	
	Between				0.133	0.069	-0.002	0.268	
	RandomEffects								

		Model 0			Model 1				
	Parameter	**Estimate**	**Standard Error**	**95% CI**	**Estimate**	**Standard Error**	**95% CI**		
	Intercept (between sibling group variability)	0.560	0.051		0.553	0.051	0.454	0.653	***

*p<0.05;
p<0.01;*p<0.001

Multilevel models of standardized test scores

Preliminary multilevel analysis using a single imputed data set

Preliminary model building and assessment was performed on a single imputed data set. The addition of level 2 variables (race/ethnicity, nativity) resulted in an 'infinite likelihood' notice. The community variables for percent poverty in census tract and percent persons African American in census tract were similar to individual level variables already in the substantive model, so these were excluded from models. Even after removal of these variables 'infinite likelihood' notifications persisted when testing models over all imputed datasets. As a result, final models were computed using select imputations for ELA (n=8) and Math (n=12) based on convergence. Appendix E summarizes the imputations for which an infinite likelihood message was received for ELA and Math models after the inclusion of race/ethnicity and nativity.

Correlations

Pairwise correlations between covariates included in the substantive model are presented in Appendix D. The correlation estimates are combined across valid imputations to yield an average correlation coefficient for each pair of variables. Most notably, the correlation coefficient between birth year and birth order was very high [r(se)= 0.72(0.01)]. As birth order has been found in studies to be strong independently associated with intellectual performance (106) and is of interest here, it will be included in the substantive models. Additionally, preliminary assessment showed no association between birth year and ELA and Math z-scores. Lastly, models including both variables resulted in extremely large intercept values that reduced when birth year was removed from the model.

English Language Arts (ELA) z-score models

Null Model

The null model estimates the within and between sibling group variance prior to the addition of other covariates. The between variance was 0.411 [p< .0001] suggesting that there is

significant sibling group-to-sibling group variance in the average ELA z-scores. The estimated within sibling group variance was 0.592 [p< .0001]. The change in these estimates from the null model to the models with gestational age and other predictors can be used to assess how much additional within and between sibling group variability can be explained by the added within and between cluster covariates. The ICC was .41 suggesting that about 41% of the variation between ELA z-scores is due to sibling group, thus supporting the need to use a multilevel analysis for these data. The AIC (mean) was 86876.50.

Model 1

Model 1 includes within and between sibling group variables for gestational age. The between sibling group variance was 0.410, meaning that compared to the null model, gestational age explained an additional .24% of the between sibling group variation in ELA z-scores. The within sibling group variance was 0.592, which indicates that compared to the null model gestational age did not explain any additional within sibling group variation in ELA z-scores. The between and within sibling group effect estimates of gestational age were 0.080 (se=0.021) and 0.014 (se=0.015), respectively. The between sibling group gestational age estimate suggests that a change increase in the mean gestational age between sibling groups is associated with a significant increase in the mean ELA z-score by 0.080 units (95% CI: 0.038, 0.121). A non-significant finding is observed for the within sibling group effect of gestational age. The results suggest that on average, a within sibling group change in gestational age (i.e. an individual sibling's deviation from the sibling group's average gestational age) is associated with a nonsignificant increase in that sibling's ELA z-score by 0.01 units (95% CI:-0.02,0.04).

Model 2

In addition to gestational age, model 2 contained the following level-1 (individual sibling level) predictors: child's sex assigned a birth, birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous

blood-lead level test, child eligibility for school lunch program, and log odds propensity score rank for gestational age at delivery.

The between sibling group variance was 0.309, meaning that compared to the null model, the addition of these predictors explained an additional 24.7% of the between sibling group variation in ELA z-scores. The within sibling group variance was 0.578, which indicates that these predictors explained an additional 2.42% of the within sibling group variation in ELA z-scores compared to the null model.

Gestational age did not explain differences in ELA z-score both within and between sibling groups. Within sibling group, a change in gestational age resulted in a nonsignificant decrease in ELA z-score (β= -0.003, 95%CI: -0.032, 0.026). Between sibling groups, a change in the average gestational age resulted in a nonsignificant increase in average ELA z-score by 0.019 units (β= 0.019, 95%CI: -0.020, 0.057).

Assessing the between sibling group effect of the remaining covariates, a change (in the mean) between sibling groups for the following predictors was associated with a significant change in the average ELA z-score: child's sex assigned a birth, birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous blood-lead level test, and child eligibility for school lunch program. Specifically, the between sibling group effect of child's sex assigned at birth is estimated as 0.238 (95% CI: 0.203,0.272). Therefore, the ELA z-score for females versus males will differ (be higher) by 0.238 units on average. The between sibling group effect of birth order is estimated as -0.300 (95% CI: -0.358, -0.241). This indicates that the ELA z-score for sibling groups with children born later in the birth order will on average be lower by 0.300 units. Difference in method of delivery between sibling groups was associated with a significant change in the average ELA z-score by 0.072 units (β= 0.072, 95%CI: 0.034, 0.111). Specifically, sibling groups with c-section births on average had a higher average ELA z-score. Similarly, a change in the mean Apgar score between sibling groups was associated with a change increase in mean ELA z-score by 0.083

units (β= 0.083, 95%CI: 0.054, 0.113). Between sibling groups, there was on average, a significant decrease in the ELA z-score associated with having had an indication requiring NICU admission. (β=-0.258, 95%CI:-0.340,-0.177). Sibling groups with small for gestational age siblings had on average a significant decrease in average ELA z-score compared to other sibling groups (β=-0.137, 95%CI:-0.184, -0.090). A change increase in the mean blood lead level (BLL) between sibling groups was associated with a significant decrease in the average ELA z-score by 0.033 units (β= -0.033, 95%CI:-0.038, -0.028). Finally, between sibling group poverty differences (as measured by school lunch eligibility) were associated with a significant decrease in average ELA z-score (β =-0.785, 95%CI:-0.825, -0.746). Thus, sibling groups with higher poverty on average had lower average ELA z-scores.

Within sibling groups, similar but slightly attenuated findings are observed. A change in sex, birth order, size for gestational age, and venous BLL resulted in a significant change in an individual sibling's ELA z-score. Specifically, being female was associated with a significant within sibling group increase in ELA z-score (β =0.165, 95%CI: 0.141,0.188). Conversely, being born later in the birth order (β= -0.094, 95%CI: -0.118, -0.070), being small for gestational age (β=-0.04729, 95% CI: -0.088,-0.007), and having a higher BLL (β=-0.010, 95%CI: -0.014,-0.006) were all associated with a significant decrease in an individual sibling's ELA z-score. Unlike between sibling groups, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, and child eligibility for school lunch program did not result in a significant change in an individual sibling's ELA z-score within a sibling group, suggesting perhaps that some shared family exposures may be explaining the relationship.

Model 3

Model 3 adds a random slope component for gestational age in order to assess if the within sibling group effect of gestational age varies across sibling groups. The explained variance changed only slightly in this model. The between sibling group variance was 0.311, meaning that compared to the null model, this model explained 24.3% of the between sibling group variation in

ELA z-scores. The within sibling group variance was 0.574, which indicates that this model explained 3.04% of the within sibling group variation in ELA z-scores compared to the null model. The random effect for gestational age was not significant (β=0.023, 95% CI: -0.031,0.076), suggesting that there was not significant variation in the slope of the within sibling group effect across sibling groups.

Within sibling groups, a change in sex, birth order, size for gestational age, and venous blood lead level (BLL) continued to result in a significant change in an individual sibling's ELA z-score. Also similar to model 2, between sibling group change in the average for child's sex assigned a birth, birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous blood-lead level test, child eligibility for school lunch program, and log odds propensity score rank for gestational age at delivery resulted in a change in the sibling group average ELA z-score.

Model 4

Model 4 included a covariate for maternal race/ethnicity combined with maternal nativity (U.S born, non US born), and dropped the insignificant random slope from model 3. In test models, the race/ethnicity, nativity composite variable explained more of the between sibling group variability than including separate covariates for race/ethnicity and nativity (41.61% versus 40.88%, respectively).

The between sibling group variance was 0.240 compared to the null model, this model explained 41.6% of the between sibling group variation in ELA z-scores. Thus, the addition of the level 2 variable race/ethnicity, nativity did much to explain the variation in ELA z-scores between sibling groups. The within sibling group variance was 0.578, which indicates that this model explained 2.36% of the within sibling group variation in ELA z-scores compared to the null model. The individual characteristics significant in previous models continued to be significant predictors for ELA z-scores within and between sibling groups; gestational age continued to show no significant impact on ELA z-scores neither within (β= -0.003, 95% CI: -0.032, 0.026) nor

between (β= 0.006565, 95%CI: -0.030, 0.043) sibling groups. The sibling group characteristic, race/ethnicity, nativity was significantly associated with changes in ELA z-score. Specifically, sibling groups of black, non-Latino US born (β= -0.667, 95%CI: -0.712, -0.622) and black, non-Latino non US born (β= -0.371, 95%CI: -0.421, -0.321) or Latino US born (β= -0.559, 95%CI: -0.604, -0.515) and Latino non US born (β= -0.412, 95%CI: -0.456, -0.367) identities on average experienced a significant decrease in average ELA z-scores compared to sibling groups of white, non-Latino, US born identity. Conversely, sibling groups of Asian/Pacific Islander, non US born identity (β= 0.177, 95%CI: 0.129, 0.225) saw an increased ELA z-score on average compared to sibling groups of white, non-Latino, US born identity. Asian/Pacific Islander, US born sibling groups had a non-significant increase in ELA z-score (β= 0.115, 95%CI: -0.075, 0.304) on average versus sibling groups of white, non-Latino, US born identity.

Model 5

Model 5 added interaction terms to test whether variability existed in the effect of gestational age on ELA z-scores by race/ethnicity, nativity. As in model 4, the between sibling group variance was 0.240, meaning that compared to the null model, this model explained 41.6% of the between sibling group variation in ELA z-scores. The within sibling group variance was 0.578, which indicates that this model explained 2.36% of the within sibling group variation in ELA z-scores compared to the null model. Thus, no additional variance was explained within or between sibling groups by the addition of the interaction terms. The individual characteristics significant in previous models were also significant predictors for ELA z-scores in this model, and gestational age continued to show no significant impact on ELA z-scores neither within nor between sibling groups. Interactions of gestational age with race/ethnicity, nativity was not significant for any race/ethnicity, nativity combination. Thus, there is no significant variation in the effect of gestational age on ELA z-scores by race/ethnicity, nativity.

Math z-score models

Null Model

The null model estimates the within and between sibling group variance prior to the addition of other covariates. The between variance was 0.375 [p< .0001] suggesting that there is significant sibling group-to-sibling group variance in the average Math z-scores. The estimated within sibling group variance was 0.594 [p< .0001]. The change in these estimates from the null model to the models with gestational age and other predictors can be used to assess how much additional within and between sibling group variability can be explained by the added within and between cluster covariates. The ICC was .387 suggesting that about 39% of the variation in Math z-scores is due to sibling group, thus supporting the need to use a multilevel analysis for these data. The AIC (mean) was 86110.68.

Model 1

Model 1 includes within and between sibling group variables for gestational age. The between sibling group variance was 0.374, meaning that compared to the null model, gestational age explained an additional .27% of the between sibling group variation in Math z-scores. The within sibling group variance was 0.594, which indicates that compared to the null model gestational age did not explain any additional within sibling group variation in Math z-scores. The between and within sibling group effect estimates of gestational age were 0.069 (se=0.021) and 0.015 (se=0.015), respectively. The between sibling group gestational age estimate suggests that a change increase in the mean gestational age between sibling groups is associated with a significant increase in the mean Math z-score by 0.069 units (95% CI: 0.029, 0.109). A non-significant finding is observed for the within sibling group effect of gestational age. The results suggest that on average, a within sibling group change in gestational age (i.e. an individual sibling's deviation from the sibling group's average gestational age) is associated with a nonsignificant increase in that sibling's Math z-score by 0.015 units (95% CI:-0.015, 0.044).

Model 2

In addition to gestational age, model 2 contained the following level-1 (individual sibling level) predictors: child's sex assigned a birth, birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous blood-lead level test, child eligibility for school lunch program, and log odds propensity score rank for gestational age at delivery.

The between sibling group variance was 0.307, meaning that compared to the null model, the addition of these predictors explained an additional 18.13% of the between sibling group variation in Math z-scores. The within sibling group variance was 0.590, which indicates that these predictors explained an additional 0.62% of the within sibling group variation in Math z-scores compared to the null model.

Gestational age did not explain differences in Math z-score both within and between sibling groups. Within sibling group, a change in gestational age resulted in a nonsignificant increase in Math z-score (β=0.007, 95%CI: -0.023, 0.036). Between sibling groups, a change in the average gestational age resulted in a nonsignificant increase in average Math z-score by 0.017 units (β= 0.017, 95%CI: -0.0218, 0.0552).

Assessing the between sibling group effect of the remaining covariates, a change (in the mean) between sibling groups for the following predictors was associated with a significant change in the average Math z-score: birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous blood-lead level test, child eligibility for school lunch program and the log odds propensity score rank of gestational age. Specifically, the between sibling group effect of birth order is estimated as -0.239 (95% CI: -0.297, -0.180). This indicates that the Math z-score for sibling groups with children born later in the birth order will on average be lower by 0.239 units. Difference in method of delivery between sibling groups was associated with a significant change in the average Math z-score by 0.042 units (β= 0.042, 95%CI: 0.004,0.080). Specifically, sibling groups with c-section births on average had a higher average Math z-score. Similarly, a change in the

mean Apgar score between sibling groups was associated with a change increase in mean Math z-score by 0.073 units (β= 0.073, 95%CI: 0.044, 0.103). Between sibling groups, there was on average, a significant decrease in the Math z-score associated with having had an indication requiring NICU admission. (β= -0.250, 95%CI: -0.320, -0.180). Sibling groups with small for gestational age siblings had on average a significant decrease in average Math z-score compared to other sibling groups (β=-0.147, 95%CI: -0.194,-0.100). A change increase in the mean blood lead level (BLL) between sibling groups was associated with a significant decrease in the average Math z-score by 0.030 units (β= -0.030, 95%CI: -0.034,-0.025). Between sibling group poverty differences (as measured by school lunch eligibility) were associated with a significant decrease in average Math z-score (β =-0.623, 95%CI: -0.662, -0.582). Thus, sibling groups with higher poverty on average had lower average Math z-scores. Lastly, a change in the average propensity score rank between sibling groups was associated with a significant increase in average Math z-score. In other words, those sibling groups with an on average higher rank for birth at 39-41 weeks also had a higher average Math z-score (β = 0.043, 95%CI: 0.029, 0.056).

Within sibling groups, similar but slightly attenuated findings are observed. A change in birth order, size for gestational age, venous BLL and log odds propensity score rank resulted in a significant change in an individual sibling's Math z-score. Being born later in the birth order (β= -0.046, 95%CI: -0.070, -0.022), being small for gestational age (β=-0.045, 95% CI: -0.086,-0.004), and having a higher BLL (β=-0.007, 95%CI: -0.011,-0.002) were all associated with a significant decrease in an individual sibling's Math z-score. Unlike between sibling groups, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, and child eligibility for school lunch program did not result in a significant change in an individual sibling's Math z-score within a sibling group. Notably and in contrast with findings for ELA z-scores, child's sex assigned at birth did not explain differences in Math z-score both within and between sibling groups. Within sibling group, females had a nonsignificant decrease in Math z-

score (β= -0.003, 95%CI: -0.027, 0.021). Between sibling groups, female sibling groups had a nonsignificant increase in average Math z-score by 0.025 units (β= 0.025, 95%CI: -0.010, 0.060).

Model 3

Model 3 adds a random slope component for gestational age in order to assess if the within sibling group effect of gestational age varies across sibling groups. The explained variance changed only slightly in this model. The between sibling group variance was 0.310, meaning that compared to the null model, this model explained 17.3% of the between sibling group variation in Math z-scores. The within sibling group variance was 0.584, which indicates that this model explained 1.63% of the within sibling group variation in Math z-scores compared to the null model. The random effect for gestational age was not significant (β=0.036, 95% CI: -0.018,0.091), suggesting that there was not significant variation in the slope of the within sibling group effect across sibling groups.

Within sibling groups, a change in birth order, size for gestational age, and venous blood lead level (BLL) continued to result in a significant change in an individual sibling's Math z-score. Also similar to model 2, between sibling group change in the average birth order, method of delivery, 5-minute Apgar score, neonatal intensive care unit (NICU) admission, small for gestational age, highest venous blood-lead level test, child eligibility for school lunch program, and log odds propensity score rank for gestational age at delivery resulted in a change in the sibling group average Math z-score.

Model 4

Model 4 included a covariate for maternal race/ethnicity combined with maternal nativity (U.S born, non US born), and dropped the insignificant random slope from model 3. In test models, the race/ethnicity, nativity composite variable explained more of the between sibling group variability than including separate covariates for race/ethnicity and nativity (42.13% versus 41.60%, respectively).

The between sibling group variance was 0.217, meaning that compared to the null model, this model explained 42.1% of the between sibling group variation in Math z-scores. As with the ELA z-score, the addition of the level 2 variable race/ethnicity, nativity helped explain a strong proportion of the variation in Math z-scores between sibling groups. The within sibling group variance was 0.590, which indicates that this model explained 0.62% of the within sibling group variation in Math z-scores compared to the null model. The individual characteristics significant in previous models continued to be significant predictors for Math z-scores within and between sibling groups; gestational age continued to show no significant impact on Math z-scores neither within (β= 0.007, 95% CI: -0.023, 0.036) nor between (β= -0.002, 95%CI: -0.037, 0.034) sibling groups. The sibling group characteristic, race/ethnicity, nativity was significantly associated with changes in Math z-score. Specifically, sibling groups of black, non-Latino US born (β= -0.630, 95%CI: -0.674, -0.585) and black, non-Latino non US born (β= -0.310, 95%CI: -0.359, -0.261) or Latino US born (β= -0.484, 95%CI: -0.528, -0.440) and Latino non US born (β= -0.258, 95%CI: -0.302, -0.214) identities on average experienced a significant decrease in average Math z-scores compared to sibling groups of white, non-Latino, US born identity. Conversely, sibling groups of Asian/Pacific Islander, non US born identity (β= 0.404, 95%CI: 0.357, 0.451) saw an increased Math z-score on average compared to sibling groups of white, non-Latino, US born identity. Asian/Pacific Islander, US born sibling groups also had a significant increase in Math z-score (β= 0.223, 95%CI: 037, 0.410) on average versus sibling groups of white, non-Latino, US born identity.

Model 5

Model 5 added interaction terms to test whether variability existed in the effect of gestational age on Math z-scores by race/ethnicity, nativity. Like model 4, the between sibling group variance was 0.217, meaning that compared to the null model, this model explained 41.6% of the between sibling group variation in Math z-scores. The within sibling group variance was 0.589, which indicates that this model explained 0.79% of the within sibling group variation in

Math z-scores compared to the null model. Thus, no additional variance was explained between sibling groups by the addition of the interaction terms, and only a very slightly higher proportion of within sibling group variance was explained by adding in of interaction terms. The individual characteristics significant in previous models were also significant predictors for Math z-scores in this model, and gestational age continued to show no significant impact on Math z-scores neither within nor between sibling groups. Interactions of gestational age with race/ethnicity, nativity was not significant for any race/ethnicity, nativity combination. Thus, there is no significant variation in the effect of gestational age on Math z-scores by race/ethnicity, nativity.

Table 4.11 Random Effects Predictors of ELA z-score among Study Sample Births-Estimates Over 8 Imputations (n=31,647)

	Model 0			Model 1			Model 2			
Parameter	**Estimate**	**Standard Error**		**Estimate**	**Standard Error**		**Estimate**	**Standard Error**	**95% CI**	
RandomEffects										
Intercept										
(between sibling group variability)	0.411	0.009	***	0.410	0.009	***	0.309	0.007	0.295 0.324	***
Residual										
(within sibling group variability)	0.592	0.007	***	0.592	0.007	***	0.578	0.006	0.566 0.591	***
Gestational age										
Variance explained[a]										
Between sibling group variability (%)	0			(0.24)			(24.82)			
Within sibling group variability (%)	0			0.00			(2.36)			
Fit Statistics										
AIC	86876.5			86865.48			83924.2			
-2 Log L	86870.5			86855.48			83878.2			

	Model 3				Model 4				Model 5			
Parameter	**Estimate**	**Standard Error**	**95% CI**	**p**	**Estimate**	**Standard Error**	**95% CI**		**Estimate**	**Standard Error**	**95% CI**	
RandomEffects												
Intercept												
(between sibling group variability)	0.311	0.008	0.296 0.326	***	0.240	0.007	0.227	0.253 ***	0.240	0.007	0.227	0.253 ***
Residual												
(within sibling group variability)	0.574	0.008	0.559 0.590	***	0.578	0.006	0.565	0.591 ***	0.578	0.006	0.565	0.590 ***
Gestational age	0.023	0.027	-0.031 0.076									
Variance explained[a]												
Between sibling group variability (%)	(24.33)				(41.61)				(41.61)			
Within sibling group variability (%)	(3.04)				(2.36)				(2.36)			
Fit Statistics												
AIC	83925.5				82007.16				82036.34			
-2 Log L	83877.5				81943.16				81936.34			

*p<0.05; **p<0.01;***p<0.001

[a] In comparison with Model 0

Table 4.12 Random Effects Predictors of Math z-score among Study Sample Births- Estimates Over 12 Imputations

	Model 0			Model 1			Model 2			
		Standard			Standard			Standard		
Parameter	**Estimate**	**Error**		**Estimate**	**Error**		**Estimate**	**Error**	**95% CI**	
Random Effects										
Between sibling group variability	0.375	0.008	***	0.374	0.008	***	0.307	0.008	0.293 0.322	***
Within sibling group variability	0.594	0.007	***	0.594	0.007	***	0.590	0.007	0.577 0.603	***
Gestational age										
Variance explained[a]										
Between sibling group variability(%)	0			(0.27)			(18.13)			
Within sibling group variability(%)	0			0.00			(0.62)			
Fit Statistics										
AIC	86110.68			86102.5			84348.36			
-2 Log L	86104.68			86092.5			84302.36			

	Model 3				Model 4				Model 5			
		Standard				Standard				Standard		
Parameter	**Estimate**	**Error**	**95% CI**		**Estimate**	**Error**	**95% CI**		**Estimate**	**Error**	**95%CI**	
Random Effects												
Between sibling group variability	0.310	0.008	0.295 0.326	***	0.217	0.007	0.205 0.230	***	0.217	0.007	0.205 0.23	***
Within sibling group variability	0.584	0.008	0.569 0.599	***	0.590	0.007	0.577 0.602	***	0.589	0.007	0.577 0.602	***
Gestational age	0.036	0.028	-0.018 0.091									
Variance explained[a]												
Between sibling group variability(%)	(17.33)				(42.13)				(42.13)			
Within sibling group variability(%)	(1.63)				(0.62)				(0.79)			
Fit Statistics												
AIC	84348.6				81827.55				81856.24			
-2 Log L	84300.6				81763.55				81756.24			

*p<0.05; **p<0.01;***p<0.001

[a]In comparison with Model 0

Table 4.11a Fixed Effects Predictors of ELA z-score among Study Sample Births- Estimates Over 8 Imputations (n=31,647)

		Model 0					Model 1					Model 2					Model 3				
			Standard					Standard					Standard					Standard			
	Parameter	Estimate	Error	95% CI		p	Estimate	Error	95% CI		p	Estimate	Error	95% CI		p	Estimate	Error	95% CI		p
	Fixed Effects																				
	Intercept	0.003	0.007	-0.011	0.016		-0.058	0.017	-0.092	-0.024		0.343	0.149	0.051	0.635	*	0.343	0.149	0.051	0.635	*
Gestational Age	Between						0.080	0.021	0.038	0.121	***	0.019	0.020	-0.020	0.057		0.019	0.020	-0.020	0.057	
	Within						0.013	0.015	-0.016	0.043		-0.003	0.015	-0.032	0.026		-0.003	0.015	-0.032	0.027	
Sex	Between											0.238	0.018	0.203	0.272	***	0.238	0.018	0.203	0.272	***
	Within											0.165	0.012	0.141	0.188	***	0.165	0.012	0.141	0.188	***
Birth Order	Between											-0.300	0.030	-0.358	-0.241	***	-0.300	0.030	-0.358	-0.242	***
	Within											-0.094	0.012	-0.118	-0.070	***	-0.094	0.012	-0.118	-0.070	***
Delivery Method	Between											0.072	0.020	0.034	0.111	***	0.072	0.020	0.034	0.111	***
	Within											0.017	0.026	-0.035	0.068		0.016	0.026	-0.035	0.068	
Apgar score, 5 minutes	Between											0.083	0.015	0.054	0.113	***	0.083	0.015	0.054	0.113	***
	Within											-0.005	0.011	-0.027	0.017		-0.005	0.011	-0.027	0.017	
NICU admission	Between											-0.258	0.040	-0.340	-0.177	***	-0.258	0.040	-0.340	-0.177	***
	Within											-0.024	0.024	-0.072	0.024		-0.024	0.024	-0.072	0.024	
Highest venous Blood	Between											-0.033	0.002	-0.038	-0.028	***	-0.033	0.002	-0.038	-0.028	***
Lead Level (bll)	Within											-0.010	0.002	-0.014	-0.006	***	-0.010	0.002	-0.014	-0.006	***
Small for Gestational Age	Between											-0.137	0.024	-0.184	-0.091	***	-0.137	0.024	-0.184	-0.091	***
	Within											-0.047	0.021	-0.088	-0.007	*	-0.047	0.021	-0.088	-0.007	*
Ever school lunch eligible	Between											-0.785	0.020	-0.825	-0.746	***	-0.785	0.020	-0.825	-0.746	***
	Within											0.002	0.029	-0.055	0.060		0.002	0.029	-0.055	0.060	
Log Odds Propensity	Between											0.049	0.007	0.034	0.063	***	0.049	0.007	0.034	0.063	***
Score Rank	Within											0.014	0.007	-0.001	0.028		0.014	0.007	-0.001	0.028	

*p<0.05; **p<0.01;***p<0.001

		Model 4					Model 5				
	Parameter	Estimate	Standard Error	95% CI			Estimate	Standard Error	95% CI		
	Fixed Effects										
	Intercept	0.560	0.140	0.285	0.835	***	0.548	0.144	0.265	0.831	***
Gestational Age	Between	0.007	0.018	-0.030	0.043		0.021	0.049	-0.074	0.116	
	Within	-0.003	0.015	-0.032	0.026		0.033	0.042	-0.050	0.116	
Sex	Between	0.236	0.017	0.204	0.268	***	0.236	0.017	0.203	0.268	***
	Within	0.165	0.012	0.141	0.188	***	0.165	0.012	0.141	0.188	***
Birth Order	Between	-0.222	0.028	-0.276	-0.167	***	-0.221	0.028	-0.276	-0.167	***
	Within	-0.094	0.012	-0.118	-0.070	***	-0.094	0.012	-0.118	-0.070	***
Delivery Method	Between	0.054	0.018	0.018	0.090	**	0.054	0.018	0.018	0.090	**
	Within	0.017	0.026	-0.035	0.068		0.016	0.026	-0.036	0.067	
Apgar score, 5 minutes	Between	0.051	0.014	0.024	0.079	***	0.052	0.014	0.024	0.079	***
	Within	-0.005	0.011	-0.027	0.017		-0.005	0.011	-0.027	0.017	
NICU admission	Between	-0.143	0.037	-0.217	-0.068	***	-0.143	0.037	-0.217	-0.068	***
	Within	-0.024	0.024	-0.072	0.024		-0.024	0.024	-0.072	0.024	
Highest venous Blood	Between	-0.022	0.002	-0.027	-0.018	***	-0.022	0.002	-0.027	-0.018	***
Lead Level (bll)	Within	-0.010	0.002	-0.014	-0.006	***	-0.010	0.002	-0.014	-0.006	***
Small for Gestational Age	Between	-0.102	0.022	-0.146	-0.058	***	-0.101	0.022	-0.145	-0.057	***
	Within	-0.047	0.021	-0.088	-0.007	*	-0.047	0.021	-0.087	-0.006	*
Ever school lunch eligible	Between	-0.501	0.021	-0.543	-0.460	***	-0.502	0.021	-0.544	-0.460	***
	Within	0.002	0.029	-0.055	0.060		0.002	0.029	-0.055	0.060	
Log Odds Propensity	Between	0.045	0.007	0.031	0.059	***	0.045	0.007	0.031	0.059	***
Score Rank	Within	0.014	0.007	-0.001	0.028		0.014	0.007	0.000	0.028	
	black,NL-NonUS	-0.371	0.026	-0.421	-0.321	***	-0.378	0.061	-0.497	-0.259	***
	black,NL-US	-0.667	0.023	-0.712	-0.622	***	-0.656	0.052	-0.758	-0.553	***
	white,NL-NonUS	-0.055	0.031	-0.115	0.005		-0.004	0.083	-0.166	0.159	
	white,NL-US		*ref*					*ref*			
Race/Ethnicity, Nativity	Latino,NonUS	-0.412	0.023	-0.456	-0.367	***	-0.399	0.054	-0.505	-0.292	***
	Latino,US	-0.559	0.023	-0.604	-0.515	***	-0.546	0.054	-0.652	-0.440	***
	Asian/PI,NonUS	0.177	0.025	0.129	0.225	***	0.194	0.061	0.075	0.314	***
	Asian/PI,US	0.115	0.097	-0.075	0.304		0.462	0.246	-0.020	0.944	
	Other,NonUS	-0.136	0.069	-0.271	0.000		-0.161	0.196	-0.546	0.223	
	Other,US	0.237	0.207	-0.169	0.643		0.179	2.021	-3.782	4.140	
B_ga_sp*RaceNat	black,NL-NonUS						0.012	0.073	-0.132	0.156	
B_ga_sp*RaceNat	black,NL-US						-0.014	0.063	-0.137	0.109	
B_ga_sp*RaceNat	white,NL-NonUS						-0.063	0.098	-0.255	0.129	
B_ga_sp*RaceNat	white,NL-US							*ref*			
B_ga_sp*RaceNat	Latino,NonUS						-0.016	0.064	-0.142	0.110	
B_ga_sp*RaceNat	Latino,US						-0.017	0.065	-0.144	0.110	
B_ga_sp*RaceNat	Asian/PI,NonUS						-0.021	0.073	-0.165	0.123	
B_ga_sp*RaceNat	Asian/PI,US						-0.456	0.297	-1.039	0.127	
B_ga_sp*RaceNat	Other,NonUS						0.035	0.233	-0.421	0.492	
B_ga_sp*RaceNat	Other,US						0.066	2.078	-4.008	4.140	
W_ga_sp*RaceNat	black,NL-NonUS						-0.032	0.061	-0.152	0.088	
W_ga_sp*RaceNat	black,NL-US						-0.072	0.053	-0.176	0.031	
W_ga_sp*RaceNat	white,NL-NonUS						-0.030	0.081	-0.189	0.128	
W_ga_sp*RaceNat	white,NL-US							*ref*			
W_ga_sp*RaceNat	Latino,NonUS						-0.036	0.054	-0.142	0.071	
W_ga_sp*RaceNat	Latino,US						-0.014	0.054	-0.120	0.091	
W_ga_sp*RaceNat	Asian/PI,NonUS						-0.047	0.061	-0.166	0.073	
W_ga_sp*RaceNat	Asian/PI,US						-0.152	0.257	-0.655	0.352	
W_ga_sp*RaceNat	Other,NonUS						-0.071	0.177	-0.417	0.276	
W_ga_sp*RaceNat	Other,US						-0.969	0.933	-2.797	0.859	

Table 4.12a. Fixed Effects Predictors of Math z-score, Estimates Over 12 Imputations (n=31,647)

		Model 0					Model 1					Model 2					Model 3				
			Standard					Standar					Standard					Standard			
	Parameter	Estimate	Error	95% CI			Estimate	d Error	95% CI			Estimate	Error	95% CI			Estimate	Error	95% CI		
	Fixed Effects																				
	Intercept	0.071	0.007	0.058	0.084	***	0.019	0.017	-0.014	0.052		0.381	0.150	0.088	0.675	*	0.381	0.150	0.088	0.674	*
Gestational Age	Between						0.069	0.021	0.029	0.109	***	0.017	0.020	-0.022	0.055		0.017	0.020	-0.022	0.055	
	Within						0.015	0.015	-0.015	0.044		0.007	0.015	-0.023	0.036		0.007	0.015	-0.023	0.036	
Sex	Between											0.025	0.018	-0.010	0.060		0.025	0.018	-0.010	0.060	
	Within											-0.003	0.012	-0.027	0.021		-0.003	0.012	-0.027	0.021	
Birth Order	Between											-0.239	0.030	-0.297	-0.180	***	-0.239	0.030	-0.297	-0.180	***
	Within											-0.046	0.012	-0.070	-0.022	***	-0.046	0.012	-0.070	-0.023	***
Delivery Method	Between											0.042	0.020	0.004	0.080	*	0.042	0.020	0.004	0.080	*
	Within											0.009	0.026	-0.042	0.060		0.009	0.026	-0.042	0.060	
Apgar score, 5 minutes	Between											0.073	0.015	0.044	0.103	***	0.073	0.015	0.044	0.103	***
	Within											0.004	0.011	-0.018	0.026		0.004	0.011	-0.018	0.026	
NICU admission	Between											-0.250	0.035	-0.320	-0.180	***	-0.250	0.035	-0.320	-0.180	***
	Within											-0.038	0.025	-0.087	0.011		-0.038	0.025	-0.087	0.012	
Highest venous Blood	Between											-0.030	0.002	-0.034	-0.025	***	-0.030	0.002	-0.034	-0.025	***
Lead Level	Within											-0.007	0.002	-0.011	-0.002	**	-0.007	0.002	-0.011	-0.002	**
Small for Gestational Age	Between											-0.147	0.024	-0.194	-0.100	***	-0.147	0.024	-0.194	-0.100	***
	Within											-0.045	0.021	-0.086	-0.004	*	-0.045	0.021	-0.086	-0.004	*
Ever school lunch eligible	Between											-0.623	0.020	-0.662	-0.583	***	-0.623	0.020	-0.662	-0.583	***
	Within											0.030	0.030	-0.028	0.088		0.031	0.030	-0.027	0.089	
Log Odds Propensity	Between											0.043	0.007	0.029	0.056	***	0.043	0.007	0.029	0.056	***
Score Rank	Within											0.021	0.007	0.008	0.035	**	0.021	0.007	0.008	0.034	**

*p<0.05; **p<0.01;***p<0.001

		Model 4					Model 5				
	Parameter	Estimate	Standard Error	95%CI			Estimate	Standard Error	95%CI		
	Fixed Effects										
	Intercept	0.550	0.138	0.280	0.821	***	0.546	0.142	0.268	0.824	***
Gestational Age	Between	-0.002	0.018	-0.037	0.034		0.008	0.048	-0.085	0.102	
	Within	0.007	0.015	-0.023	0.036		0.004	0.043	-0.080	0.087	
Sex	Between	0.024	0.016	-0.008	0.056		0.024	0.016	-0.008	0.056	
	Within	-0.003	0.012	-0.027	0.021		-0.003	0.012	-0.027	0.021	
Birth Order	Between	-0.148	0.027	-0.202	-0.095	***	-0.149	0.027	-0.202	-0.095	***
	Within	-0.046	0.012	-0.070	-0.022	***	-0.046	0.012	-0.070	-0.023	***
Delivery Method	Between	0.029	0.018	-0.007	0.064		0.028	0.018	-0.007	0.064	
	Within	0.009	0.026	-0.042	0.060		0.008	0.026	-0.043	0.059	
Apgar score, 5 minutes	Between	0.038	0.014	0.011	0.065	**	0.038	0.014	0.011	0.065	**
	Within	0.004	0.011	-0.018	0.026		0.004	0.011	-0.018	0.026	
NICU admission	Between	-0.124	0.032	-0.188	-0.060	***	-0.125	0.032	-0.189	-0.061	***
	Within	-0.038	0.025	-0.087	0.011		-0.038	0.025	-0.088	0.011	
Highest venous Blood Lead Level	Between	-0.018	0.002	-0.022	-0.013	***	-0.018	0.002	-0.022	-0.013	***
	Within	-0.007	0.002	-0.011	-0.002	**	-0.007	0.002	-0.011	-0.002	**
Small for Gestational Age	Between	-0.110	0.022	-0.153	-0.067	***	-0.110	0.022	-0.153	-0.067	***
	Within	-0.045	0.021	-0.086	-0.004	*	-0.045	0.021	-0.086	-0.004	*
Ever school lunch eligible	Between	-0.375	0.021	-0.416	-0.334	***	-0.375	0.021	-0.416	-0.334	***
	Within	0.030	0.030	-0.028	0.088		0.030	0.030	-0.028	0.088	
Log Odds Propensity Score Rank	Between	0.036	0.006	0.024	0.049	***	0.036	0.006	0.024	0.049	***
	Within	0.021	0.007	0.008	0.035	**	0.021	0.007	0.008	0.034	**
Race/Ethnicity, Nativity	black,NL-NonUS	-0.310	0.025	-0.359	-0.261	***	-0.312	0.060	-0.430	-0.195	***
	black,NL-US	-0.630	0.023	-0.674	-0.585	***	-0.604	0.051	-0.705	-0.503	***
	white,NL-NonUS	0.056	0.030	-0.003	0.115		0.076	0.082	-0.084	0.236	
	white,NL-US	0.000		*ref*			0.000		*ref*		
	Latino,NonUS	-0.258	0.022	-0.302	-0.214	***	-0.246	0.053	-0.351	-0.142	***
	Latino,US	-0.484	0.023	-0.528	-0.440	***	-0.494	0.053	-0.598	-0.391	***
	Asian/PI,NonUS	0.404	0.024	0.357	0.451	***	0.416	0.060	0.298	0.533	***
	Asian/PI,US	0.223	0.095	0.037	0.410	*	0.166	0.242	-0.308	0.641	
	Other,NonUS	-0.035	0.068	-0.169	0.098		-0.049	0.193	-0.427	0.329	
	Other,US	0.395	0.203	-0.003	0.794		0.500	1.980	-3.381	4.381	
B_ga_sp*RaceNat	black,NL-NonUS						0.003	0.072	-0.139	0.144	
B_ga_sp*RaceNat	black,NL-US						-0.035	0.062	-0.156	0.087	
B_ga_sp*RaceNat	white,NL-NonUS						-0.026	0.096	-0.215	0.164	
B_ga_sp*RaceNat	white,NL-US						0.000		*ref*		
B_ga_sp*RaceNat	Latino,NonUS						-0.015	0.063	-0.139	0.109	
B_ga_sp*RaceNat	Latino,US						0.014	0.064	-0.111	0.139	
B_ga_sp*RaceNat	Asian/PI,NonUS						-0.015	0.072	-0.157	0.126	
B_ga_sp*RaceNat	Asian/PI,US						0.075	0.293	-0.499	0.649	
B_ga_sp*RaceNat	Other,NonUS						0.017	0.229	-0.433	0.466	
B_ga_sp*RaceNat	Other,US						-0.110	2.037	-4.103	3.883	
W_ga_sp*RaceNat	black,NL-NonUS						0.024	0.062	-0.097	0.145	
W_ga_sp*RaceNat	black,NL-US						-0.036	0.053	-0.140	0.069	
W_ga_sp*RaceNat	white,NL-NonUS						-0.032	0.082	-0.192	0.128	
W_ga_sp*RaceNat	white,NL-US						0.000		*ref*		
W_ga_sp*RaceNat	Latino,NonUS						-0.032	0.055	-0.140	0.076	
W_ga_sp*RaceNat	Latino,US						0.045	0.054	-0.062	0.151	
W_ga_sp*RaceNat	Asian/PI,NonUS						0.061	0.062	-0.059	0.182	
W_ga_sp*RaceNat	Asian/PI,US						-0.091	0.259	-0.599	0.418	
W_ga_sp*RaceNat	Other,NonUS						-0.026	0.178	-0.375	0.324	
W_ga_sp*RaceNat	Other,US						0.006	0.942	-1.840	1.852	

Sensitivity Analysis

The results of various sensitivity analyses aimed at testing the robustness of estimates, are described below.

Complete Case Analysis

In models that used the original data set that included observations with missing values on some covariates (valid n=11,640) , a within cluster deviation from the family average gestational age for an individual sibling resulted in a significant decrease in ELA z-score (β= -0.065, 95%CI= -0.118, -0.012). A non-significant change decrease in Math z-score was observed for within sibling group change in gestational age (β= -0.039, 95%CI= -0.091, 0.014). Additionally, within cluster ELA z-scores were impacted by birth order, child's sex assigned at birth and BLL. Notably, there were no other variables found to significantly impact individual ELA z-score within sibling group. Within sibling group, birth order, BLL, and size for gestational age were found to have a significant impact on an individual sibling's Math z-score. Between sibling groups, a change in the average gestational age resulted in non-significant increases in average ELA (β= 0.030, 95%CI= -0.024, 0.085) and Math z-scores (β= 0.052, 95%CI= -0.002, 0.106).

Table 4.13 Multilevel Predictors of ELA z-score among Study Sample Births - Complete Case Analysis (n=11640)

	Parameter	Estimate	Standard Error	95% CI		
	Fixed Effects					
	Intercept	0.364	0.203	-0.035	0.762	
Gestational Age	Between	0.030	0.028	-0.024	0.085	
	Within	-0.065	0.027	-0.118	-0.012	*
Sex	Between	0.229	0.025	0.180	0.278	***
	Within	0.197	0.021	0.155	0.239	***
Birth Order	Between	-0.222	0.042	-0.305	-0.139	***
	Within	-0.105	0.021	-0.147	-0.064	***
Delivery Method	Between	0.052	0.027	0.000	0.105	*
	Within	0.020	0.048	-0.074	0.114	
Apgar score, 5 minutes	Between	0.069	0.021	0.029	0.109	***
	Within	-0.005	0.019	-0.043	0.032	
NICU admission	Between	-0.149	0.051	-0.248	-0.050	**
	Within	-0.069	0.055	-0.177	0.038	
Highest venous Blood Lead Level	Between	-0.024	0.003	-0.030	-0.018	***
	Within	-0.011	0.004	-0.018	-0.003	**
Small for Gestational Age	Between	-0.067	0.036	-0.137	0.003	
	Within	-0.056	0.039	-0.133	0.020	
Ever school lunch eligible	Between	-0.443	0.029	-0.500	-0.387	***
	Within	-0.013	0.047	-0.105	0.079	
Log Odds Propensity Score Rank	Between	0.044	0.008	0.028	0.060	***
	Within	0.011	0.013	-0.013	0.036	
Race/Ethnicity, Nativity	Asian/PI,NonUS	0.192	0.034	0.126	0.258	***
	Asian/PI,US	0.118	0.134	-0.144	0.381	
	Latino,NonUS	-0.374	0.032	-0.437	-0.312	***
	Latino,US	-0.546	0.032	-0.609	-0.483	***
	Other,NonUS	-0.178	0.093	-0.360	0.005	
	Other,US	0.243	0.271	-0.288	0.774	
	black,NL-NonUS	-0.360	0.037	-0.432	-0.289	***
	black,NL-US	-0.590	0.034	-0.657	-0.522	***
	white,NL-NonUS	-0.044	0.043	-0.129	0.040	
	white,NL-US		*ref*			
	Random Effects					
	Intercept	0.237	0.015			***
	Residual	0.569	0.015			***

*p<0.05; **p<0.01;***p<0.001

Table 4.13a Multilevel Predictors of Math z-score among Study Sample Births - Complete Case Analysis (n=11640)

	Parameter	Estimate	Standard Error	95% CI		
Fixed Effects						
	Intercept	0.237	0.202	-0.159	0.633	
Gestational Age	Between	0.052	0.028	-0.002	0.106	
	Within	-0.039	0.027	-0.091	0.014	
Sex	Between	0.003	0.025	-0.046	0.052	
	Within	0.010	0.021	-0.031	0.052	
Birth Order	Between	-0.161	0.042	-0.244	-0.079	***
	Within	-0.051	0.021	-0.092	-0.010	*
Delivery Method	Between	0.023	0.026	-0.029	0.075	
	Within	0.012	0.047	-0.081	0.105	
Apgar score, 5 minutes	Between	0.069	0.020	0.029	0.109	***
	Within	-0.013	0.019	-0.050	0.024	
NICU admission	Between	-0.170	0.050	-0.268	-0.071	***
	Within	-0.071	0.054	-0.178	0.036	
Highest venous Blood Lead Level	Between	-0.016	0.003	-0.022	-0.010	***
	Within	-0.008	0.004	-0.015	0.000	*
Small for Gestational Age	Between	-0.090	0.035	-0.159	-0.021	*
	Within	-0.082	0.039	-0.158	-0.006	*
Ever school lunch eligible	Between	-0.325	0.029	-0.381	-0.269	***
	Within	0.057	0.047	-0.035	0.148	
Log Odds Propensity Score Rank	Between	0.037	0.008	0.021	0.053	***
	Within	0.007	0.012	-0.017	0.032	
Race/Ethnicity, Nativity	Asian/PI,NonUS	0.419	0.034	0.353	0.485	***
	Asian/PI,US	0.010	0.133	-0.252	0.271	
	Latino,NonUS	-0.238	0.032	-0.300	-0.177	***
	Latino,US	-0.465	0.032	-0.527	-0.402	***

Parameter	Estimate	Standard Error	95% CI		
Other,NonUS	-0.029	0.093	-0.210	0.153	
Other,US	0.405	0.269	-0.123	0.933	
black,NL-NonUS	-0.303	0.036	-0.374	-0.231	***
black,NL-US	-0.561	0.034	-0.628	-0.494	***
white,NL-NonUS	0.032	0.043	-0.053	0.116	
white,NL-US		*ref*			
Random Effects					
Intercept	0.239	0.015			***
Residual	0.557	0.015			***

*p<0.05; **p<0.01;***p<0.001

Mother's age >= 40 at 1st birth

In models that included siblings where the mother's age at first live birth was over 39 (n=188), a within cluster deviation from the family average gestational age for an individual sibling resulted in a non-significant increase in ELA (β= 0.13, 95%CI= -0.17, 0.43) and Math z-score (β= 0.24, 95%CI= -0.11, 0.59). Notably, within cluster ELA z-scores were impacted by birth order (β= -0.28, 95%CI= -0.50,-0.05), such that the later in the birth order an individual sibling was born, their individual ELA z-score decreased by .28 SD units. Notably, there were no other variables found to significantly impact individual ELA z-score within sibling group. For none of the variables examined did the deviation from the sibling group average have a significant impact on an individual sibling's Math z-score. Said differently, there were no within sibling group differences that significantly explained the variance in Math z-scores.

Table 4.14 Multilevel Predictors of ELA z-score among Study Sample Births -Mother >=age 40 at 1st birth (n=188)

	Parameter	Estimate	Standard Error	95% CI		
	Fixed Effects					
	Intercept	4.16	2.84	-1.40	9.72	
Gestational Age	Between	-0.02	0.23	-0.46	0.42	
	Within	0.13	0.15	-0.17	0.44	
Sex	Between	0.23	0.21	-0.19	0.64	
	Within	0.24	0.13	-0.02	0.49	
Birth Order	Between	-0.70	1.02	-2.71	1.31	
	Within	-0.28	0.11	-0.50	-0.05	*
Delivery Method	Between	-0.40	0.19	-0.77	-0.03	*
	Within	-0.32	0.27	-0.85	0.22	
Apgar score, 5 minutes	Between	-0.25	0.22	-0.68	0.18	
	Within	0.21	0.14	-0.06	0.49	
NICU admission	Between	-0.35	0.53	-1.44	0.74	
	Within	0.02	0.33	-0.64	0.68	
Highest venous Blood Lead Level	Between	-0.02	0.03	-0.08	0.04	
	Within	0.00	0.02	-0.04	0.04	
Small for Gestational Age	Between	-0.08	0.31	-0.69	0.53	
	Within	0.19	0.22	-0.25	0.63	
Ever school lunch eligible	Between	-0.11	0.25	-0.61	0.39	
	Within	0.16	0.26	-0.35	0.66	
Log Odds Propensity Score Rank	Between	0.12	0.08	-0.04	0.27	
	Within	0.05	0.07	-0.09	0.18	
Race/Ethnicity, Nativity	black,NL-NonUS	-0.54	0.30	-1.13	0.04	
	black,NL-US	-1.01	0.35	-1.69	-0.33	**
	white,NL-NonUS	0.03	0.48	-0.92	0.97	
	white,NL-US		*ref*			
	Latino,NonUS	-0.75	0.29	-1.31	-0.18	*
	Latino,US	-0.64	0.33	-1.29	0.01	
	Asian/PI,NonUS	-0.33	0.30	-0.92	0.27	
	Asian/PI,US	0.15	0.49	-0.82	1.11	
	Other,NonUS	-0.07	0.68	-1.40	1.25	
	Random effects					
	Intercept	0.24	0.07	0.11	0.37	***
	Residual	0.35	0.05	0.25	0.45	***

*p<0.05; **p<0.01;***p<0.001

Table 4.14a Multilevel Predictors of ELA z-score among Study Sample Births -Mother >=age 40 at 1st birth (n=188)

	Parameter	Estimate	Standard Error	95% CI		
	Fixed Effects					
	Intercept	3.26	2.59	-1.82	8.34	
Gestational Age	Between	-0.11	0.21	-0.53	0.30	
	Within	0.24	0.18	-0.11	0.59	
Sex	Between	-0.18	0.20	-0.57	0.20	
	Within	0.14	0.15	-0.15	0.43	
Birth Order	Between	-0.47	0.92	-2.27	1.32	
	Within	0.00	0.13	-0.26	0.26	
Delivery Method	Between	-0.49	0.17	-0.82	-0.15	**
	Within	-0.18	0.29	-0.76	0.39	
Apgar score, 5 minutes	Between	-0.19	0.21	-0.59	0.21	
	Within	0.17	0.16	-0.14	0.48	
NICU admission	Between	-0.33	0.43	-1.17	0.51	
	Within	-0.10	0.32	-0.72	0.53	
Highest venous Blood Lead Level	Between	0.02	0.03	-0.03	0.08	
	Within	0.01	0.03	-0.04	0.06	
Small for Gestational Age	Between	-0.09	0.29	-0.66	0.48	
	Within	-0.08	0.26	-0.58	0.43	
Ever school lunch eligible	Between	-0.09	0.24	-0.55	0.38	
	Within	0.19	0.29	-0.38	0.76	
Log Odds Propensity Score Rank	Between	0.16	0.07	0.02	0.31	*
	Within	0.03	0.08	-0.13	0.19	
Race/Ethnicity, Nativity	black,NL-NonUS	-0.84	0.28	-1.38	-0.29	**
	black,NL-US	-0.97	0.32	-1.59	-0.34	**
	white,NL-NonUS	0.02	0.45	-0.86	0.91	
	white,NL-US		*ref*			
	Latino,NonUS	-0.80	0.27	-1.33	-0.27	**
	Latino,US	-0.69	0.31	-1.30	-0.08	*
	Asian/PI,NonUS	-0.36	0.28	-0.91	0.19	
	Asian/PI,US	0.48	0.46	-0.43	1.38	
	Other,NonUS	-0.12	0.63	-1.36	1.12	
	Random effects					
	Intercept	0.14	0.06	0.01	0.26	*
	Residual	0.45	0.07	0.32	0.58	***

*p<0.05; **p<0.01;***p<0.001

Fixed Effects Model

The fixed effects model, which only allows for estimation of within cluster (i.e. within sibling group) effects, confirmed findings of earlier models. A within cluster deviation from the family average gestational age for an individual sibling did not result in a significant change in ELA (β= -0.003, 95%CI= -0.03, 0.03) nor Math z-score (β= 0.007, 95%CI= -0.02, 0.04). Also similar to hybrid random effects models, the fixed effects model found that individual sibling within cluster ELA z-scores were impacted by birth order (β= -0.09, 95%CI= -0.12,-0.07), sex assigned at birth (β= 0.16, 95%CI=0.14,0.19), venous blood lead level (β= -0.01, 95%CI= -0.01,-0.01), and size for gestational age (β= -0.05, 95%CI=-0.09,-0.01). Math z-scores were impacted by birth order (β= -0.05, 95%CI= -0.07,-0.02), venous blood lead level (β= -0.01, 95%CI= -0.002,-0.01), size for gestational age (β= -0.05, 95%CI=-0.09,-0.004) and the log odds propensity score rank (β= 0.02, 95%CI=0.01,0.03).

Table 4.15 Multilevel Predictors of ELA z-score among Study Sample Births -Fixed Effects Model (n=31,647)

Parameter	**Estimate**	**Standard Error**	**95% CI**		
Gestational Age	-0.003	0.015	-0.032	0.026	
Sex	0.165	0.012	0.141	0.188	***
Birth Order	-0.094	0.012	-0.118	-0.070	***
Delivery Method	0.017	0.026	-0.035	0.068	
Apgar score, 5 minutes	-0.005	0.011	-0.027	0.017	
NICU admission	-0.024	0.024	-0.072	0.024	
Highest venous Blood Lead Level	-0.010	0.002	-0.014	-0.006	***
Small for Gestational Age	-0.047	0.021	-0.088	-0.007	*
Ever school lunch eligible	0.002	0.029	-0.055	0.060	
Log Odds Propensity Score Rank	0.014	0.007	-0.001	0.028	

*p<0.05; **p<0.01;***p<0.001

Table 4.15a Multilevel Predictors of Math z-score among Study Sample Births - Fixed Effects Model (n=31,647)

Parameter	Estimate	Standard Error	95% CI		
Gestational Age	0.007	0.015	-0.023	0.036	
Sex	-0.003	0.012	-0.027	0.021	
Birth Order	-0.046	0.012	-0.070	-0.022	***
Delivery Method	0.009	0.026	-0.042	0.060	
Apgar score, 5 minutes	0.004	0.011	-0.018	0.026	
NICU admission	-0.038	0.025	-0.087	0.011	
Highest venous Blood Lead Level	-0.007	0.002	-0.011	-0.002	**
Small for Gestational Age	-0.045	0.021	-0.086	-0.004	*
Ever school lunch eligible	0.030	0.030	-0.028	0.088	
Log Odds Propensity Score Rank	0.021	0.007	0.008	0.035	**

*p<0.05; **p<0.01;***p<0.001

Summary

Findings across models suggest that on average, a within sibling group change in gestational age is associated with a nonsignificant change in ELA z-score and Math z-score. Between sibling groups, a change in the average gestational age, was associated with a non-significant change in the average ELA and Math z-scores. These findings suggest that gestational age in the term period may not be significantly impacting school performance.

CHAPTER 5

DISCUSSION

The primary objective of this study was to understand the relationship between gestational age and standardized test scores using siblings in a multilevel analysis to explore where the sources of variation in this relationship exist.

For the first research question which sought to explore if there is a difference in school performance by gestational age among siblings born in the term period in NYC, this study found that gestational age group showed no statistically significant impact on test scores neither within nor between sibling groups. In the model that included term gestational age as the sole explanatory variable, it explained almost none of the variation. Moreover, there was no statistically significant difference found in ELA and Math z-scores by gestational age group in the term period controlling for other factors. Among children born between 37 and 41 weeks, statistically significant variations observed in standardized test scores for siblings in the same family and between children in different families were explained by factors such as sex assigned at birth, birth order, venous blood lead level (BLL) and differences in pregnancy characteristics exclusive of gestational age; specifically, being born small for gestational age (SGA). This suggests that to the extent that in utero characteristics are influencing school test scores, they may be doing so independently of gestational age when children are born between 37 and 41 weeks. Plus, the attenuation of effect estimates within sibling groups versus between sibling groups suggests unmeasured confounding may remain in the between sibling groups estimates by variables also impacting school outcomes (e.g. homework help, school attended, presence of reading material in the home, parent/guardian- child interaction).

An interpretation of effect size estimates based on educational intervention literature suggests 0.25 of a standard deviation to be a large effect; 0.15 a medium effect; and 0.05 to 0.10 a small (but non-negligible) effect. (55)

Thus, comparing 2 siblings born at identical weeks gestation in the term period, with one sibling born SGA and one born AGA, the sibling born SGA will have an ELA and a Math score on average .05 of a standard deviation unit lower than that of their AGA sibling. The difference in Math scores can be likened to the reduction in Math achievement for the average 4th grader due to 10 days of teacher absence throughout the school year. (107) The ELA and Math effect estimates approach the average effect (0.08 of a SD unit) found from participation in supplemental reading and math services for one school year. As such, the reduction in score for the sibling born SGA aligns with having nearly 1.7 to 2.4 months less classroom instruction. (55)

Additional analysis using birthweight in place of SGA as a measure of birth size ('smallness') confirmed previous significant findings. Specifically, between sibling groups, ELA z-score increased by .042 of a standard deviation unit on average (β= 0.042, 95%CI= 0.027, 0.057) and the Math z-score increased by .048 on average (β= 0.048, 95%CI=-0.033, 0.063) for every 500-gram (~1.1 lbs) increase in average sibling group birthweight; within sibling groups, an individual sibling's ELA z-score increased by .023 standard deviation units (β= 0.023, 95%CI= 0.005, 0.040) and the Math z-score increased by .028 standard deviation units (β= 0.028, 95%CI= 0.010, 0.045) on average for every 500-gram increase in birthweight. With birthweight in the model (replacing SGA), both within and between sibling groups, there was no significant difference in test scores for early term versus full term births. Test scores also varied significantly within and between sibling groups by sex assigned at birth, birth order, and BLL confirming previous results. (Appendix G)

The relationship between test scores and being born SGA has been observed in other studies, which have found this association across the gestational age spectrum; that is, regardless of gestational age, infants born SGA had poorer academic performance. (2,4,34). In addition, studies of children ages 4 to 7 who were of the same gestational age found that compared to children born appropriate for gestational age (AGA), those born SGA displayed reduced grey and white matter volumes in the areas of the brain that impact thinking, learning and speech, and

regional variations that suggest a different development of the area of the brain that controls math learning (108). SGA and/or low birth weight (both among births in the term range and preterm) may be an indication of fetal growth restriction (FGR). While not all small for gestational age or low birth weight neonates experience FGR (i.e. neonates born SGA may be 'constitutionally small' and haven't experienced any adverse in utero events restricting their growth), a strong proportion do experience it (109). On all SGA, but particularly on the FGR births it is paramount to recognize and systematically collect data on the mechanism underlying their condition. Placental insufficiency has been identified as the primary cause of FGR, and results in chronic fetal hypoxia (i.e. insufficient oxygen supply) which can have consequences for neurodevelopment. (110) Even though the fetal response to hypoxia includes redistribution of cardiac output to spare vital organs, this may not ensure normal brain development. (110) Critical to the impact is likely the timing of the onset of FGR vis-à-vis sensitive periods of brain development. (110) Early onset FGR occurs prior to 28 weeks of gestation, and is usually diagnosed by the 2nd trimester. (109,110) Given this timing, it has the potential to affect neurodevelopmental processes such as the appearance of the first sulcus (13-17 weeks) through the closure of the lateral sulcus (27-28 weeks). (111) The process of sulcation is a sign of cortical maturation which is important in the thinking, learning, speech and emotional development processes. (108,111,112) Late onset FGR occurs after 28 weeks of gestation, and is more frequent than early onset (~70-80% of FGR). (109,110) Late onset FGR infants also experience fetal hypoxia resulting in adverse neurodevelopmental outcomes, and both early and late onset FGR increase the risk for cerebral palsy. (110) To the extent that routine surveillance can include collecting detailed data on the timing of onset of FGR (and potentially mechanisms involved) it may help to elucidate more data for action.

The association of test scores with sex assigned at birth (ELA only), birth order and BLL (even at low levels) affirms findings in the literature from non-sibling based studies, and, as the

associations persisted in within sibling group analyses, reinforces the validity of the results. (106,113–116)

A second study objective was to understand if a relationship between gestational age and school performance varies according to race/ethnicity. Interactions of gestational age with race/ethnicity and nativity were not significant for any combination of race/ethnicity and nativity reaffirming that differences in test scores are not the result of anything intrinsic to a particular group, but more so a reflection of the numerous unmeasured social and cultural factors that impact race/ethnicity and nativity, e.g. family structure and parenting style, structured social inequality in access to resources, neighborhood deprivation, etc. (117,118)

Strengths and Limitations

The strengths of this study lie in the diversity of the population, as well as the vast array of variables available via the data warehouse to control for confounding and extends the work of previous studies of gestational age and school standardized test scores in NYC birth populations by controlling for within-sibling group characteristics using sibling-based analysis. Educational literature on family home environment and school performance suggests a very strong influence of the home environment on school achievement (and other behaviors); thus, controlling for this factor in assessing relationships between gestational age and school outcomes is vital. (53,54,62,119) The hybrid random effects model which is used in this study simultaneously estimates within and between sibling group variations, and provides less biased within-sibling group estimates versus random effects or GEE models, while using the sibling group mean of covariates to account for between cluster variation. (78) While models could be computed without the between sibling group (i.e. cluster mean) components, excluding them would result in inadequate control for other sources of level 2 (between sibling group) variation; thus estimates of level 2 variables such as race/ethnicity, and nativity would be over- or underestimated. (80) In addition, assessing interaction by race, a consideration not previously addressed in studies of school outcomes will encourage further research into the role of the life course and structural and

social barriers in affecting birth outcomes, specifically if it results in variations in intrauterine effects (not necessarily gestational age). Control for maternal conditions and pregnancy complications using propensity scores to organize the distribution between gestational age groups provides a mechanism to address confounding by indication, as many of these conditions are medical indications for early term (or even late preterm deliveries). (120) Lastly, this study controls for BLL, which is missing from much of the literature exploring the relationship gestational age and school outcomes. Lead is a known neurotoxin, and as it likely explains some of the variation in school test scores, it should be controlled for when assessing if an independent relationship exists between gestational age and school test scores.

One important limitation of this study is representativeness. There were about 200,000 term births in NYC between 1994 and 1998 who subsequently attended public school in grade 3 this represents about 34% of the entire NYC birth population from 1994-1998. Data from the American Community Survey for 2006-2008 (with 2006 being the year that the 1998 births would be in grade 3), estimate that about 75% of NYC elementary school students attended public schools. (121) However, how this is divided by race/ethnicity is of note. Over 65% of students were Latino or black, non-Latino, while only 17% were white, non-Latino. This contrasts with the private school makeup which was approximately 64% white, non-Latino. Thus, the sample used in this study is likely reflective of the non-white population of NYC. If and how the make-up of this sample affects the results is difficult to say.

This study utilizes birth certificate and other health department data that were initially collected for surveillance and administrative purposes; as a result, there is variability in how and by whom data were collected at individual institutions for birth certificate data. This can contribute to biased estimates of effect due to over or underreporting of factors that affect gestational age at birth. Of note, when compared to the medical record or similar as the gold standard, studies have reported the sensitivity of maternal medical risk factors and comorbidities as low as 5%; in contrast, pregnancy outcomes (gestational age, gender, birth weight, Apgar

scores, and delivery method) were found to have high levels of agreement, and sensitivities that were greater than 90 percent. (122–124) Further, the whole cohort associations that remained significant while controlling for shared family characteristics (sex assigned at birth, birth order, BLL and small for gestational age) support stronger causative impact of these variables on the outcome of school performance.

Bias due to unmeasured confounders that vary within sibling groups (and therefore between sibling groups) is also a limitation for consideration. (125) Some of these factors include Adverse Childhood Experiences (ACEs); environmental toxins, and school related factors. First, while this study controls for individual-level poverty (using school lunch eligibility as a proxy), the fluidity over time and within families of other ACEs such as violence in the home, abuse or neglect, attempted or completed suicide by a parent or other family member, parental substance use disorder or mental health disorder and their impact on school outcomes are an important consideration (126) The timing of exposure to environmental chemicals/toxins vis-à-vis fetal development and the extent to which pre-pregnancy exposure to environmental toxins alters fetal developmental processes are important considerations, as many chemicals to which the population is routinely exposed via personal care products or household cleaning products can affect cognitive and behavioral development. An analysis of samples from 268 pregnant women in the National Health and Nutritional Examination Survey (NHANES) 2003-2004 revealed detectable levels of certain polychlorinated biphenyls (PCBs), organochlorine pesticides (OCs), perfluorinated compounds (PFCs), phenols and phthalates in nearly all the women included. (127) This is notable as the chemicals detected show evidence for placental and breastmilk transfer, and in particular, PFCs and PCBs carry the risk of reduced fetal growth. (128) Levels of exposure to these and other chemicals are not routinely measured nor collected during the routine course of prenatal care, thus it is unknown the extent to which these factors may contribute to differences in school outcomes. Exploring the impact of school factors such as school location, teacher quality and experience along with other community level units of classification (e.g. neighborhood

poverty) compels using a cross-classified multilevel model given that some children do not belong to only one higher level unit. (81,129–131) In this study specifically, if some children were born in different neighborhoods or attended a different school than their siblings, they belong to a different neighborhood or school cluster than others in their sibling group. The impact of not accounting for this is that some of the variation attributed to the sibling group level might be neighborhood or school level effects. In addition, individual sibling differences with regards to these variables may also further explain some of the observed within-cluster variation. The limited timeframe over which births are being assessed (1994-1998), makes the size of cross-classified clusters particularly small, which may cause estimation issues, and would impact the precision of model estimates. (81) Computing a cross-classified model should however be a consideration for a larger study with similar objectives.

In further consideration of cluster size, literature has noted that when clusters are samples of the full cluster and the cluster size is small with level 2 variables as a function of aggregated level 1 variables e.g. the average blood lead level for a sibling group, the regression estimates are underestimated. (132) A study of childlessness and family size revealed that U.S. women near the end of their childbearing years (ages 40 – 44) in 1994 had only 1 or 2 children. (133) Thus, a sibling group in this study with mother's first birth at age ≥ 40 years likely contains good representation of the full sibling group. Sensitivity analysis conducted among women with a first birth at ≥ 40 years old confirmed findings that gestational age did not impact test scores within and between sibling groups.

While a total of 25 multiply imputed data sets were created, final analyses used only the imputed datasets which achieved model convergence: n=8 for ELA and n=12 for Math. Despite this, the relative efficiency, i.e. how well the true population parameters are estimated is between 94% and 98% with at least 5 imputations when the fraction of missing information (FMI) is between 20 and 30%. (89) As the FMI existed within this range, the number of imputed data sets utilized appears adequate.

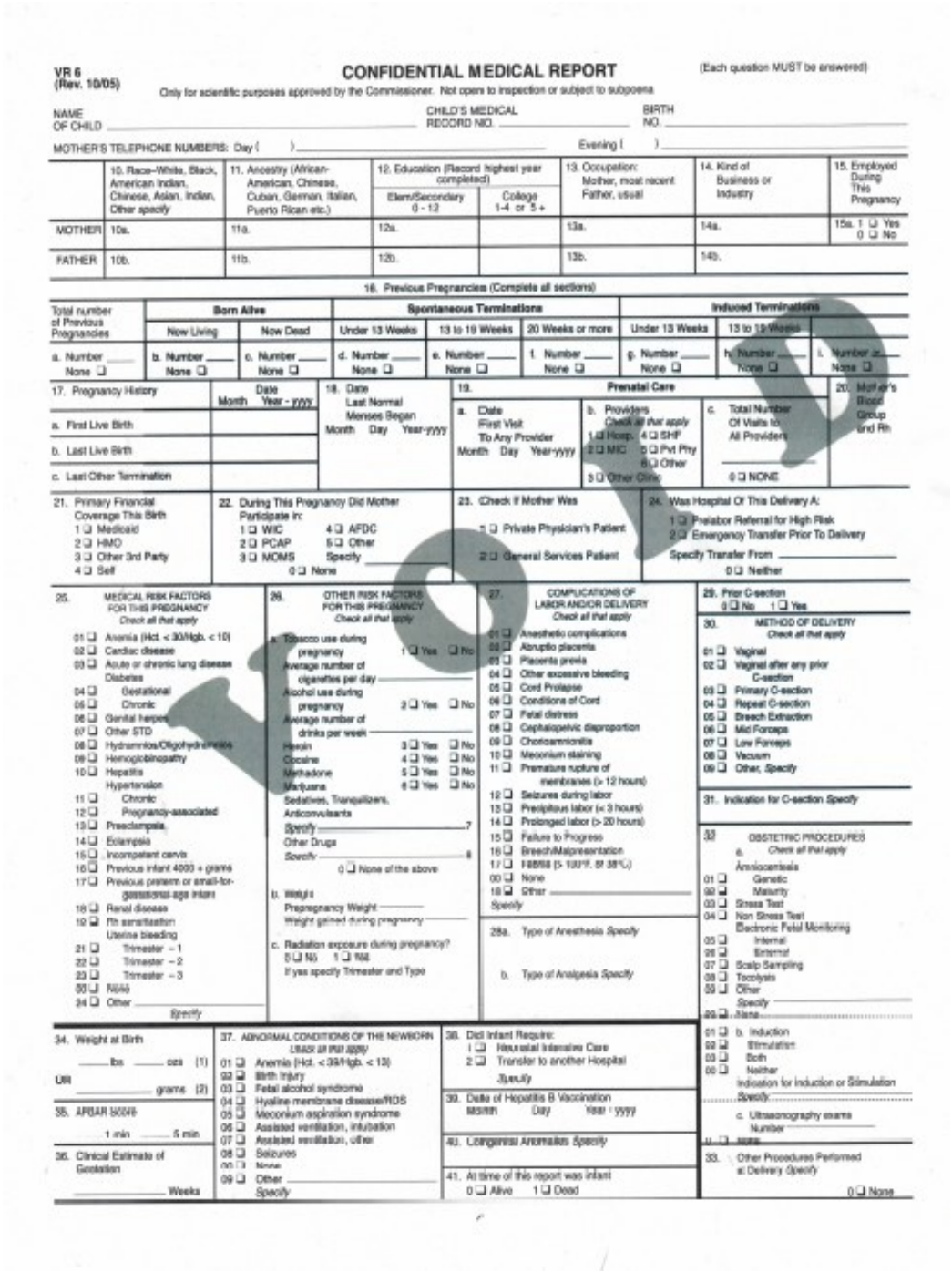
Study Implications and Recommendations

There are several considerations for clinical medicine and public health. As one example, the practice of non-medically indicated C-section deliveries prior to 39 weeks has been the focus of several reduction initiatives due to the findings of increased neonatal morbidities even among births of 37- or 38-weeks' gestation (120,134,135). Even though this study didn't find subject-specific standardized test performance to be impacted by early term delivery, there are still the proximal adverse neonatal outcomes (e.g. respiratory distress, transient tachypnea, NICU admission, and pneumonia) to consider. (120) The findings in this study also have implications for prenatal health to address proximal exposures that contribute to conditions such as small for gestational age and low birth weight (where it relates to fetal growth restriction) but more importantly to address all aspects of health over the life course. Thus, breaking down structured social inequities that influence maternal mental and physical health is essential, as the persistence of inequities into the post-natal environment especially without the buffer of protective factors can further impact a child's health and academic outcomes. Additionally, since experiences have a strong capacity to affect brain development (a substantial portion of which occurs postnatally), particularly in the early childhood years, the cumulative and dynamic effect of negative exposures such as ACEs including (but not limited to) growing up in poverty, maternal psychological distress, familial incarceration or substance use disorder may be reflected not only in poorer academic performance, but also in poorer long term outcomes. (51,53,118,126,136) Lastly, data sets using more contemporary birth cohorts are needed to evaluate the reliability of findings, especially given the changing landscape of exposures in the pre and post-natal environments, such as the introduction of more maternal health supplements targeted at improving fetal brain development (i.e. DHA supplements), and the deep challenges posed by the current opioid crisis and subsequent risk for neonatal abstinence syndrome. A local consideration in NYC surrounds the impact of implementation of Universal Pre-K (age 4) and 3-K programs on later test score performance.

Non-significant differences in test scores among children in the same family born early term versus full term indicates that other pre- and post-natal factors may be at work in impacting school success; there are still many discoveries to be made. To further elucidate root causes for child health and scholastic outcomes, consideration should be given to developing deeper and broader frameworks for data collection and on-going linkage of various surveillance systems (vital event, education, health, etc.) to support identification of risk factors over the life course for adverse maternal and child health outcomes and assess whether these factors are perpetuated or mitigated over time, within and between families.

APPENDIX A. VARIABLES SUMMARY

Figure A1. Confidential Medical Report Of Birth



VR 6
(Rev. 10/05)

CONFIDENTIAL MEDICAL REPORT

(Each question MUST be answered)

Only for scientific purposes approved by the Commissioner. Not open to inspection or subject to subpoena

NAME OF CHILD _____ CHILD'S MEDICAL RECORD NO. _____ BIRTH NO. _____

MOTHER'S TELEPHONE NUMBERS: Day ( ) _____ Evening ( ) _____

	10. Race–White, Black, American Indian, Chinese, Asian, Indian, Other *specify*	11. Ancestry (African-American, Chinese, Cuban, German, Italian, Puerto Rican etc.)	12. Education (Record highest year completed) Elem/Secondary 0 - 12	College 1-4 or 5 +	13. Occupation: Mother, most recent Father, usual	14. Kind of Business or Industry	15. Employed During This Pregnancy
MOTHER	10a.	11a.	12a.		13a.	14a.	15a. 1 ❑ Yes 0 ❑ No
FATHER	10b.	11b.	12b.		13b.	14b.	

16. Previous Pregnancies (Complete all sections)

Total number of Previous Pregnancies	Born Alive		Spontaneous Terminations			Induced Terminations		
	Now Living	Now Dead	Under 13 Weeks	13 to 19 Weeks	20 Weeks or more	Under 13 Weeks	13 to 19 Weeks	
a. Number ____ None ❑	b. Number ____ None ❑	c. Number ____ None ❑	d. Number ____ None ❑	e. Number ____ None ❑	f. Number ____ None ❑	g. Number ____ None ❑	h. Number ____ None ❑	i. Number ____ None ❑

17. Pregnancy History	Date Month Year - yyyy	18. Date Last Normal Menses Began Month Day Year-yyyy	19. Prenatal Care a. Date First Visit To Any Provider Month Day Year-yyyy	b. Providers Check all that apply 1 ❑ Hosp. 4 ❑ SHF 2 ❑ MD 5 ❑ Pvt Phy 6 ❑ Other 3 ❑ Other Clinic	c. Total Number Of Visits for All Providers 0 ❑ NONE	20. Mother's Blood Group and Rh
a. First Live Birth						
b. Last Live Birth						
c. Last Other Termination						

21. Primary Financial Coverage This Birth	22. During This Pregnancy Did Mother Participate In:	23. Check If Mother Was	24. Was Hospital Of This Delivery A:
1 ❑ Medicaid 2 ❑ HMO 3 ❑ Other 3rd Party 4 ❑ Self	1 ❑ WIC 4 ❑ AFDC 2 ❑ PCAP 5 ❑ Other 3 ❑ MOMS Specify _____ 0 ❑ None	1 ❑ Private Physician's Patient 2 ❑ General Services Patient	1 ❑ Prelabor Referral for High Risk 2 ❑ Emergency Transfer Prior To Delivery Specify Transfer From _____ 0 ❑ Neither

25. MEDICAL RISK FACTORS FOR THIS PREGNANCY *Check all that apply*

01 ❑ Anemia (Hct. < 30/Hgb. < 10)
02 ❑ Cardiac disease
03 ❑ Acute or chronic lung disease
Diabetes
04 ❑ Gestational
05 ❑ Chronic
06 ❑ Genital herpes
07 ❑ Other STD
08 ❑ Hydramnios/Oligohydramnios
09 ❑ Hemoglobinopathy
10 ❑ Hepatitis
Hypertension
11 ❑ Chronic
12 ❑ Pregnancy-associated
13 ❑ Preeclampsia
14 ❑ Eclampsia
15 ❑ Incompetent cervix
16 ❑ Previous infant 4000 + grams
17 ❑ Previous preterm or small-for-gestational-age infant
18 ❑ Renal disease
19 ❑ Rh sensitization
Uterine bleeding
21 ❑ Trimester – 1
22 ❑ Trimester – 2
23 ❑ Trimester – 3
50 ❑ None
24 ❑ Other _____ *Specify*

26. OTHER RISK FACTORS FOR THIS PREGNANCY *Check all that apply*

a. Tobacco use during pregnancy 1 ❑ Yes ❑ No
Average number of cigarettes per day _____
Alcohol use during pregnancy 2 ❑ Yes ❑ No
Average number of drinks per week _____
Heroin 3 ❑ Yes ❑ No
Cocaine 4 ❑ Yes ❑ No
Methadone 5 ❑ Yes ❑ No
Marijuana 6 ❑ Yes ❑ No
Sedatives, Tranquilizers, Anticonvulsants *Specify* _____ 7
Other Drugs *Specify* _____ 8
0 ❑ None of the above

b. Weight
Prepregnancy Weight _____
Weight gained during pregnancy _____

c. Radiation exposure during pregnancy?
0 ❑ No 1 ❑ Yes
If yes specify Trimester and Type

27. COMPLICATIONS OF LABOR AND/OR DELIVERY *Check all that apply*

01 ❑ Anesthetic complications
02 ❑ Abruptio placenta
03 ❑ Placenta previa
04 ❑ Other excessive bleeding
05 ❑ Cord Prolapse
06 ❑ Conditions of Cord
07 ❑ Fetal distress
08 ❑ Cephalopelvic disproportion
09 ❑ Chorioamnionitis
10 ❑ Meconium staining
11 ❑ Premature rupture of membranes (> 12 hours)
12 ❑ Seizures during labor
13 ❑ Precipitous labor (< 3 hours)
14 ❑ Prolonged labor (> 20 hours)
15 ❑ Failure to Progress
16 ❑ Breech/Malpresentation
17 ❑ Fibrin (> 100°F, or 38°C)
00 ❑ None
18 ❑ Other _____ *Specify*

28a. Type of Anesthesia *Specify*

b. Type of Analgesia *Specify*

29. Prior C-section 0 ❑ No 1 ❑ Yes

30. METHOD OF DELIVERY *Check all that apply*

01 ❑ Vaginal
02 ❑ Vaginal after any prior C-section
03 ❑ Primary C-section
04 ❑ Repeat C-section
05 ❑ Breech Extraction
06 ❑ Mid Forceps
07 ❑ Low Forceps
08 ❑ Vacuum
09 ❑ Other, *Specify*

31. Indication for C-section *Specify*

32. OBSTETRIC PROCEDURES
a. *Check all that apply*
Amniocentesis
01 ❑ Genetic
02 ❑ Maturity
03 ❑ Stress Test
04 ❑ Non Stress Test
Electronic Fetal Monitoring
05 ❑ Internal
06 ❑ External
07 ❑ Scalp Sampling
08 ❑ Tocolysis
09 ❑ Other
Specify _____
99 ❑ None

01 ❑ b. Induction
02 ❑ Stimulation
03 ❑ Both
00 ❑ Neither
Indication for Induction or Stimulation *Specify* _____

c. Ultrasonography exams Number _____

11 ❑ None

33. Other Procedures Performed at Delivery *Specify*
0 ❑ None

34. Weight at Birth
_____ lbs _____ ozs (1)
OR
_____ grams (2)

35. APGAR SCORE
1 min _____ 5 min

36. Clinical Estimate of Gestation
_____ Weeks

37. ABNORMAL CONDITIONS OF THE NEWBORN *Check all that apply*

01 ❑ Anemia (Hct. < 39/Hgb. < 13)
02 ❑ Birth Injury
03 ❑ Fetal alcohol syndrome
04 ❑ Hyaline membrane disease/RDS
05 ❑ Meconium aspiration syndrome
06 ❑ Assisted ventilation, intubation
07 ❑ Assisted ventilation, other
08 ❑ Seizures
09 ❑ Other _____ *Specify*

38. Did Infant Require:
1 ❑ Neonatal Intensive Care
2 ❑ Transfer to another Hospital
Specify

39. Date of Hepatitis B Vaccination
Month Day Year - yyyy

40. Congenital Anomalies *Specify*

41. At time of this report was infant
0 ❑ Alive 1 ❑ Dead

Table A1. Variables used in assessing the relationship between term gestational age and standardized test scores among siblings born in NYC to NYC resident mothers, 1994-1998

Variables	Measurement Level (Individual or Sibling Group)	Used In Propensity Model*
Mother's Race/Ethnicity	sibling group	
Mother's Age	individual	✓
Mother's Education Completed	individual	✓
Mother's Nativity	sibling group	
Father's Age	individual	✓
Father's Education Completed	individual	✓
Was child ever eligible for school lunch program	individual	
Mother Employed During This Pregnancy	individual	✓
Primary Financial Coverage for Pregnancy and Birth	individual	✓
Mother's Weight	individual	✓
Mother's tobacco use during pregnancy	individual	✓
Mother's Alcohol Use During Pregnancy	individual	✓
Mother's Use of Amphetamines, cocaine, heroin, methadone or marijuana	individual	✓
Diabetes	individual	✓
Pregnancy-related hypertension	individual	✓
Chronic hypertension	individual	✓
No prenatal care received	individual	✓
Medical Risk Factor - Previous Preterm or SGA Infant	individual	✓
Medical Risk Factor - Renal Disease	individual	✓
Other maternal risk factors	individual	✓
Complication of Labor/Delivery - Cord Prolapse	individual	✓
Complication of Labor/Delivery - Fetal Distress	individual	✓
Complication of Labor/Delivery - Abruption Placenta	individual	✓
Complication of Labor/Delivery - Placenta Previa	individual	✓
Method of Delivery	individual	
Child's year of birth	individual	
Birth Order	individual	
Child's Sex assigned a birth	individual	
Apgar Score, 5 Minutes	individual	
Any Congenital Anomaly	individual	✓
Small for Gestational Age (<10th percentile of birthweight for gestational age)	individual	

Variables	Measurement Level (Individual or Sibling Group)	Used In Propensity Model*
Neonatal Intensive Care Unit Admission	individual	
Highest venous blood-lead level test	individual	
Gestational Age at Delivery**	individual	✓
Child's earliest 3rd grade English Language Arts (ELA) z-score	individual	
Child's earliest 3rd grade Math z-score	individual	
sibling id (not analyzed)	sibling group	✓

*variables used in the propensity score model appear in test score models as a composite rank

**main exposure used as outcome in computing propensity score model

APPENDIX B. MULTIPLE IMPUTATION SUMMARY

Figure B1a-c. Trace Plots for Continuous Imputed Variables

Figure B1a. Apgar score

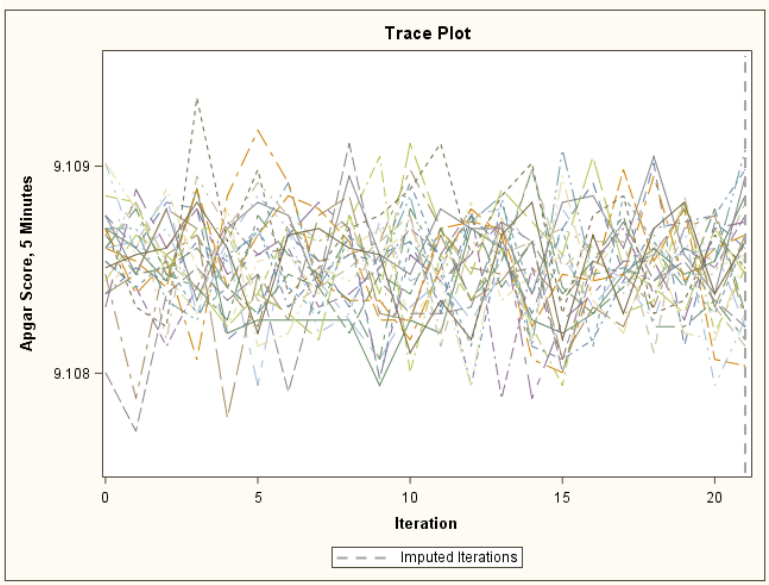


Figure B1b. Highest venous blood lead level test

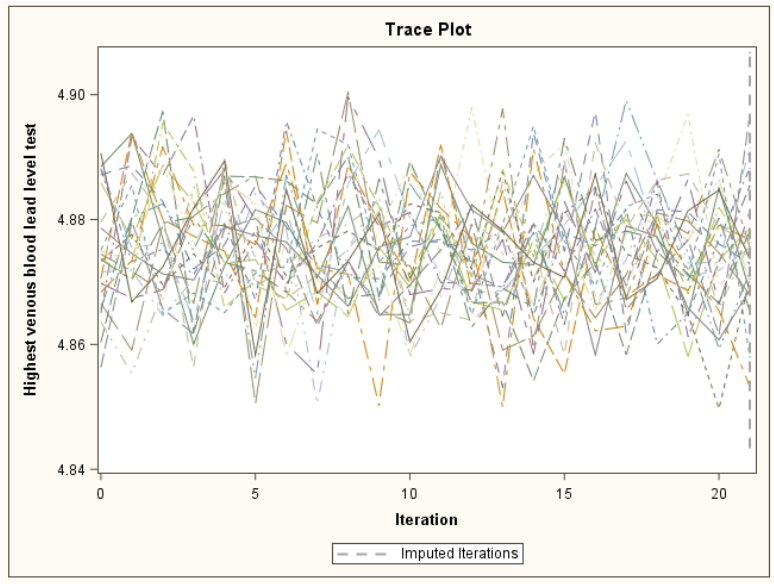


Figure B1c. Father's age

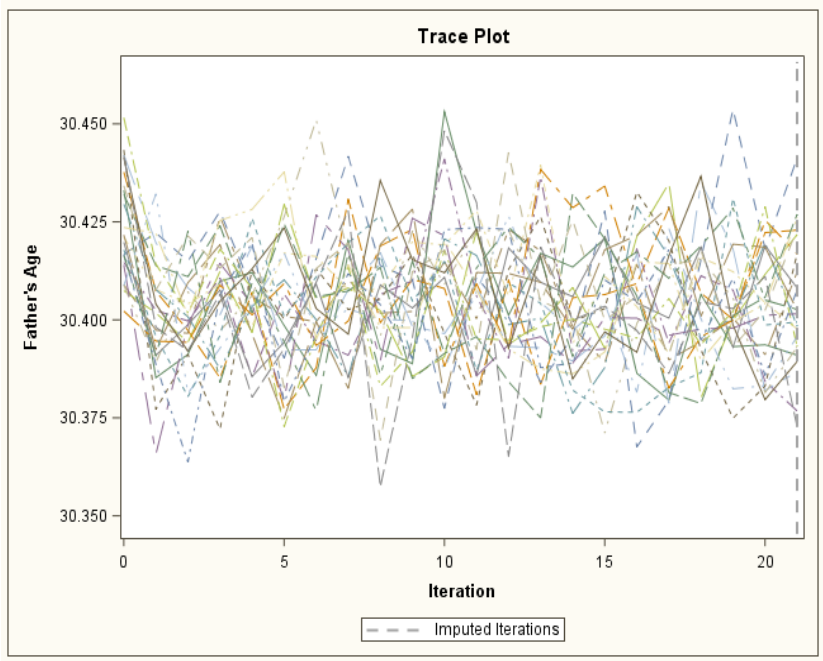
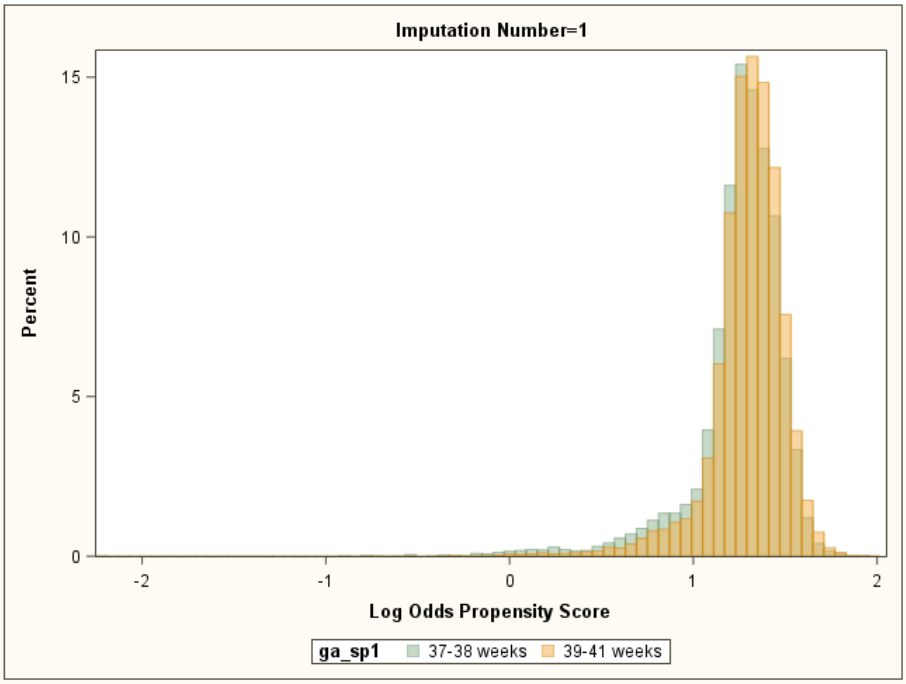

Trace Plot
Father's Age
30.450
30.425
30.400
30.375
30.350
0
5
10
15
20
Iteration
Imputed Iterations

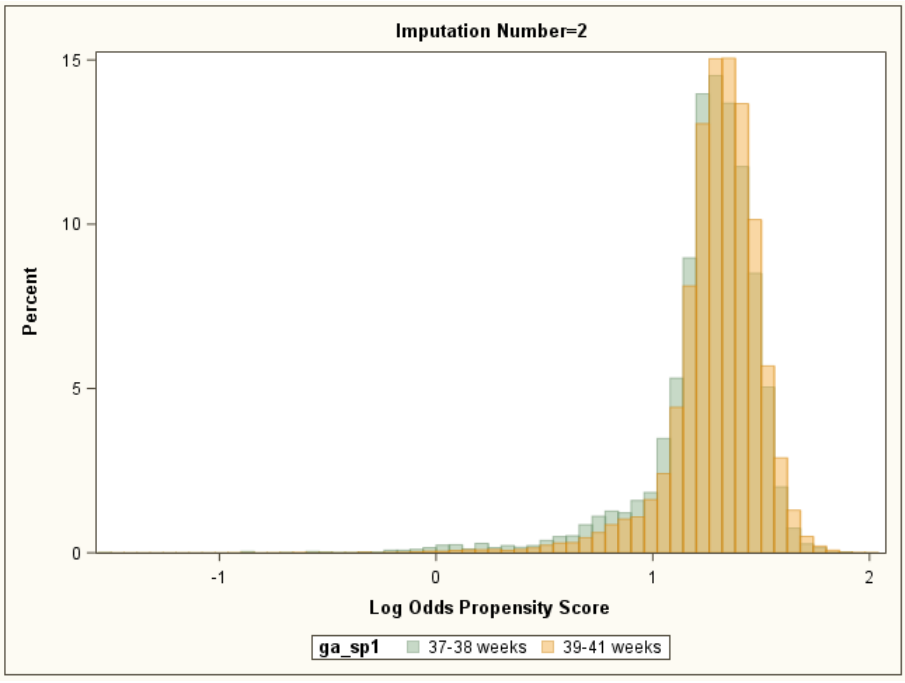
Table B1. Comparison of Imputed and Original Values for Imputed Variables

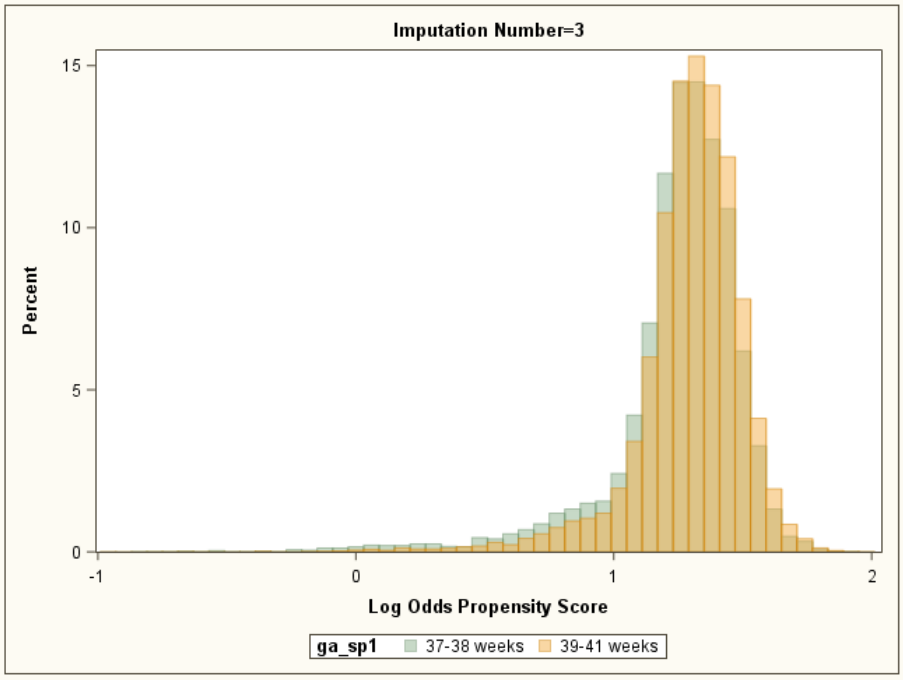
	Summary Over 25 Imputed Data Sets				Original Data			
Variable	**Mean**	**Standard Error**	**95% Confidence Limits**		**Mean**	**Standard Error**	**95% Confidence Limits**	
Medical Risk Factor - Previous Preterm or SGA Infant	0.005	0.000	0.004	0.006	0.005	0.000	0.004	0.006
Medical Risk Factor - Renal Disease	0.001	0.000	0.001	0.001	0.001	0.000	0.001	0.001
Complication of Labor/Delivery - Cord Prolapse	0.001	0.000	0.001	0.002	0.001	0.000	0.001	0.002
Complication of Labor/Delivery - Fetal Distress	0.029	0.001	0.027	0.031	0.029	0.001	0.027	0.031
Complication of Labor/Delivery - Abruption Placenta	0.003	0.000	0.002	0.003	0.003	0.000	0.002	0.003
Complication of Labor/Delivery - Placenta Previa	0.003	0.000	0.002	0.003	0.003	0.000	0.002	0.003
Apgar Score, 5 Minutes	9.109	0.003	9.102	9.115	9.109	0.003	9.102	9.115
Mother's tobacco use during pregnancy	0.064	0.001	0.061	0.066	0.063	0.001	0.060	0.066
Method of Delivery	1.160	0.002	1.156	1.164	1.160	0.002	1.156	1.164
Primary Financial Coverage for Pregnancy and Birth	1.691	0.005	1.680	1.701	1.691	0.006	1.680	1.702
Mother's Alcohol Use During Pregnancy	0.006	0.000	0.006	0.007	0.005	0.000	0.004	0.006
Mother's Education Completed	2.036	0.006	2.026	2.047	2.040	0.006	2.029	2.051
NoPrenatalCare	0.027	0.001	0.025	0.029	0.025	0.001	0.023	0.027
Mother Employed During This Pregnancy	0.252	0.003	0.247	0.257	0.251	0.003	0.246	0.256
Mother's Weight	0.076	0.002	0.073	0.080	0.077	0.002	0.074	0.081
Neonatal Intensive Care Unit Admission	0.094	0.003	0.089	0.099	0.061	0.001	0.058	0.064
Highest venous blood-lead level test	4.875	0.024	4.828	4.921	4.867	0.024	4.819	4.915
Father's Age	30.405	0.043	30.321	30.489	31.056	0.044	30.969	31.143
Father's Education Completed	2.142	0.006	2.130	2.154	2.225	0.006	2.213	2.238

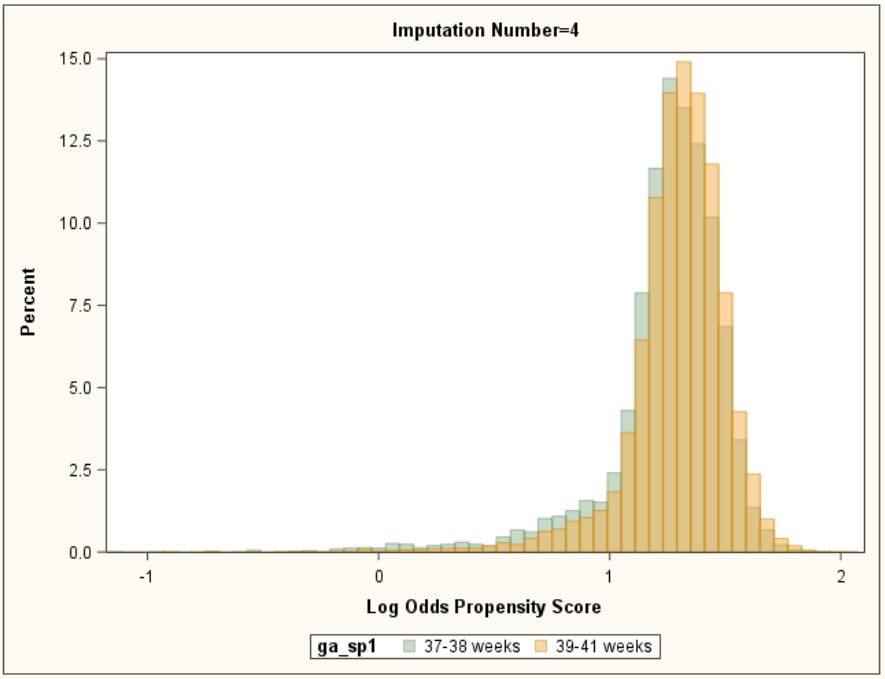
APPENDIX C. PROPENSITY SCORE PLOTS

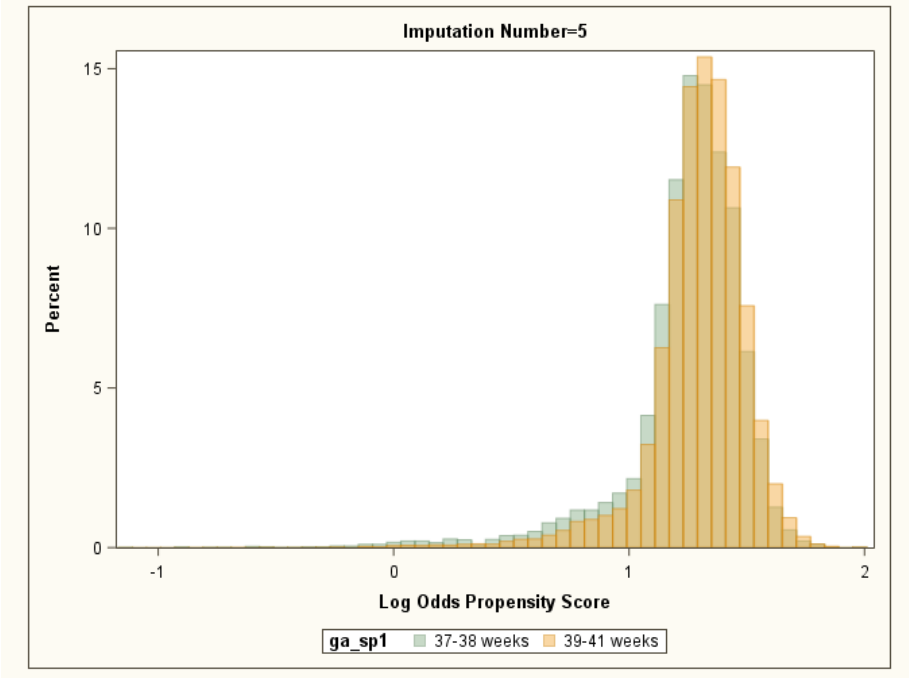
Figure C1 – C25. Plots of Overlap of Log Odds Propensity Score between Early Term and Full Term Births (25 Imputations)

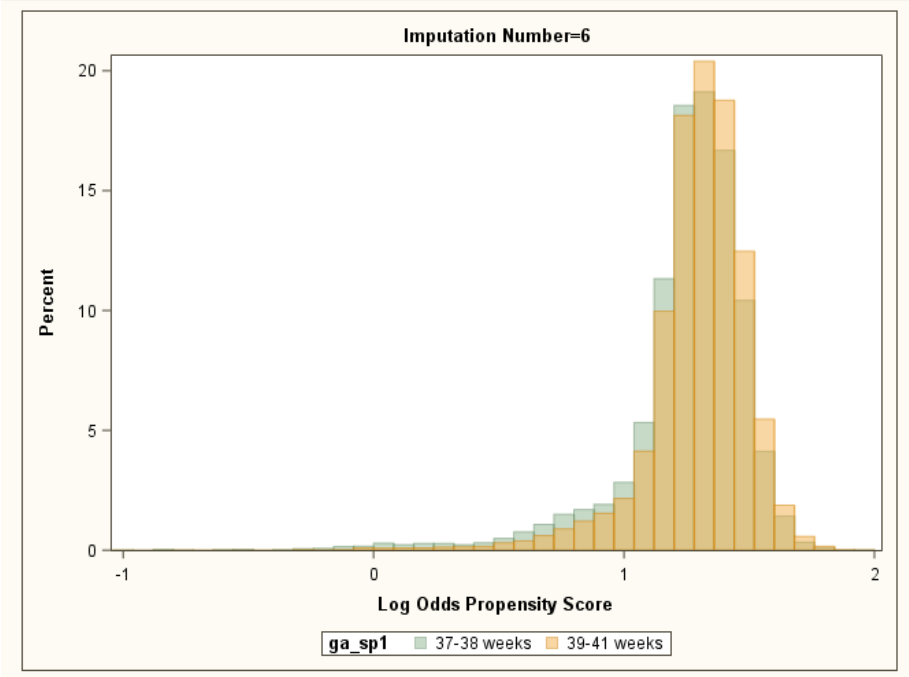


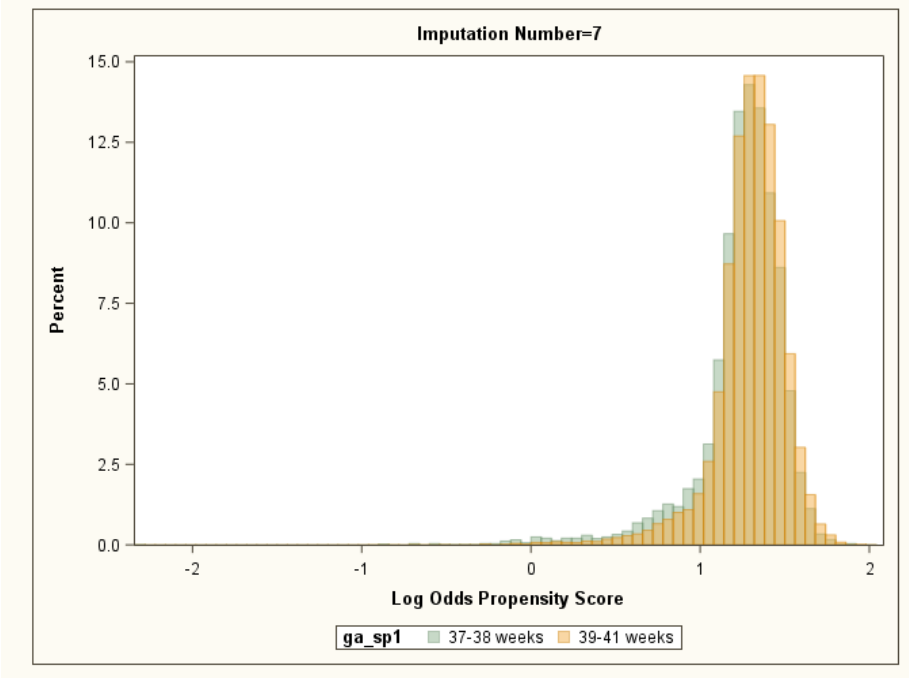


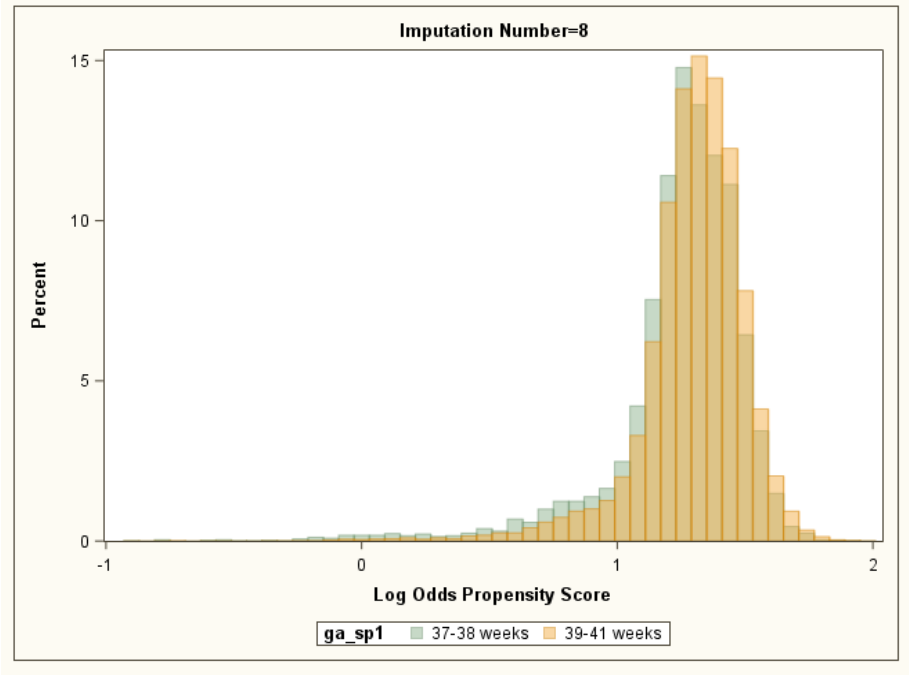

Imputation Number=3
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

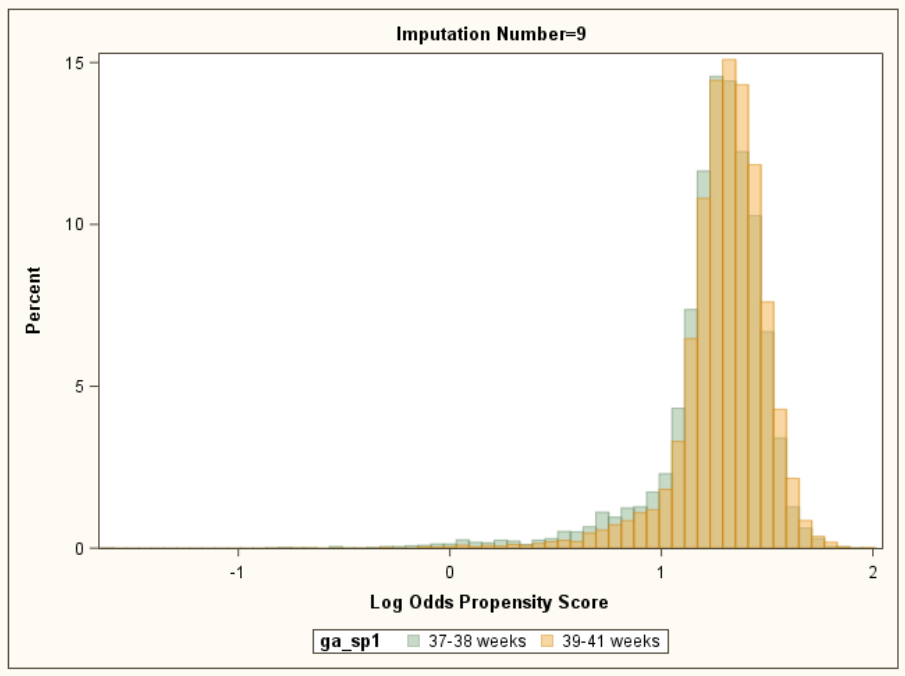

Imputation Number=4
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

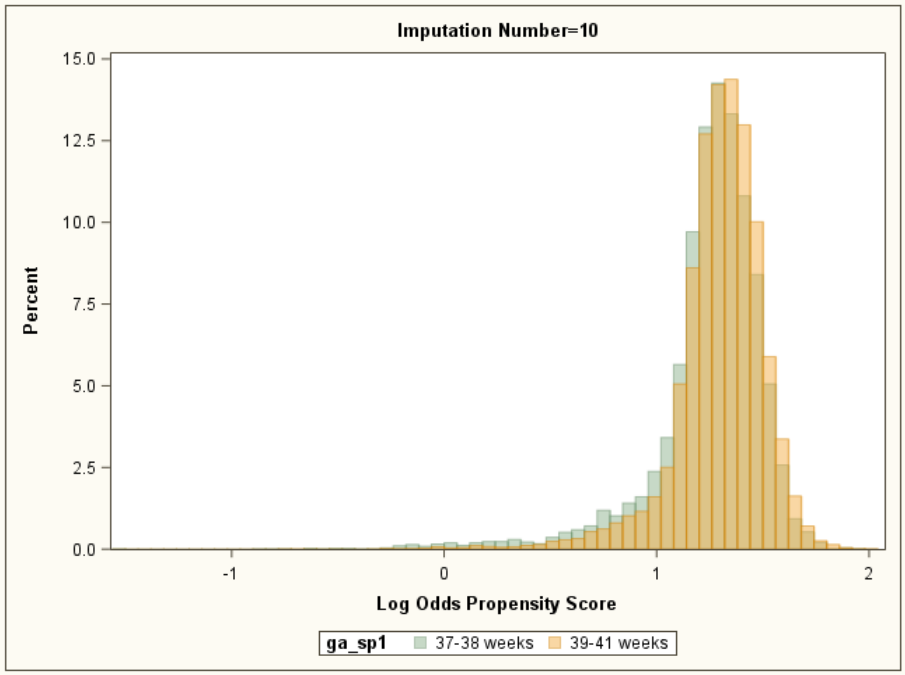

Imputation Number=5
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

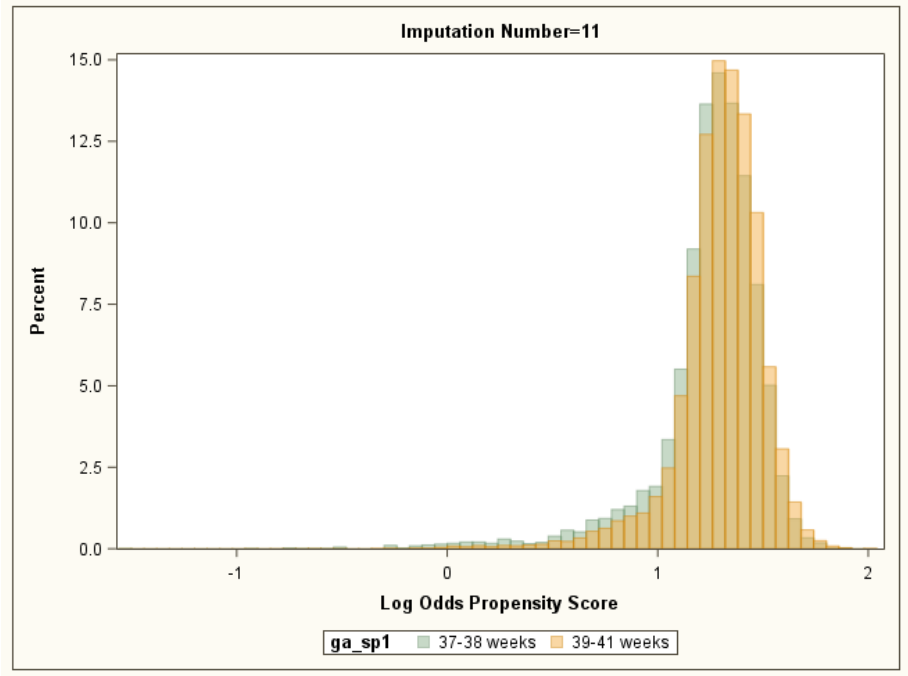

Imputation Number=6
Percent
20
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

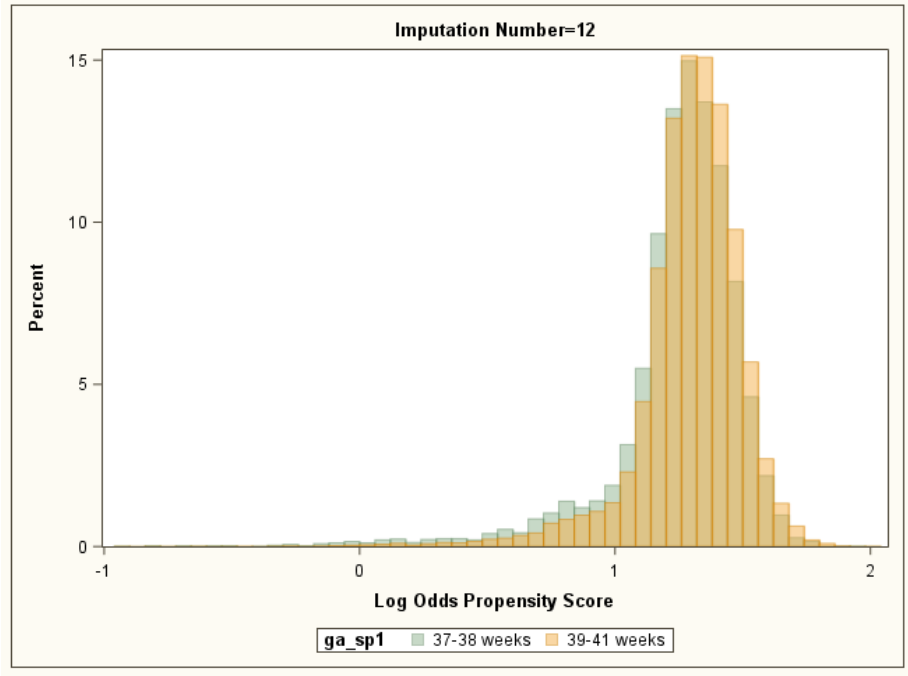

Imputation Number=7
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

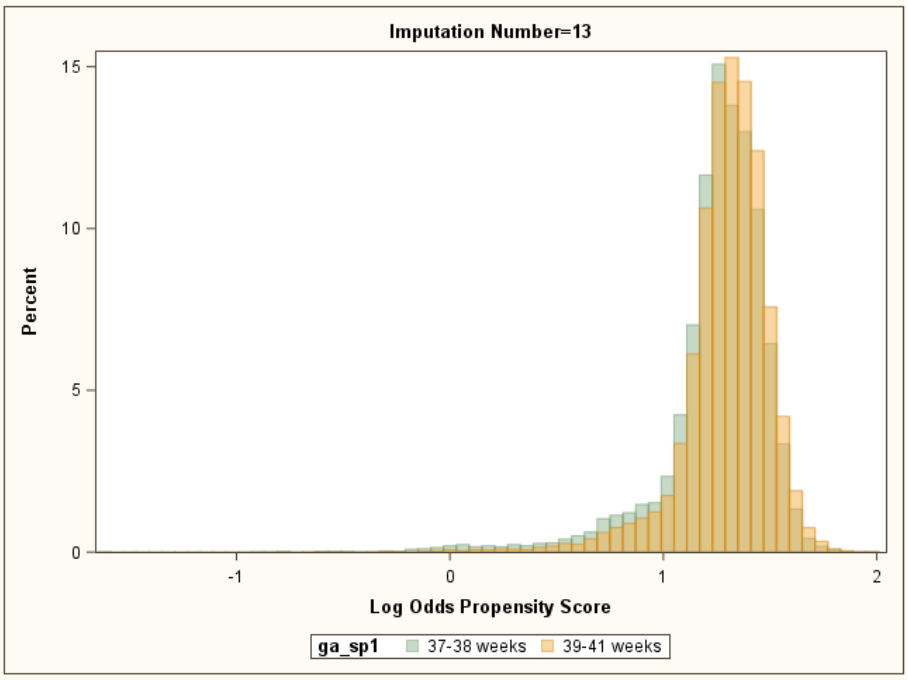

Imputation Number=8
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

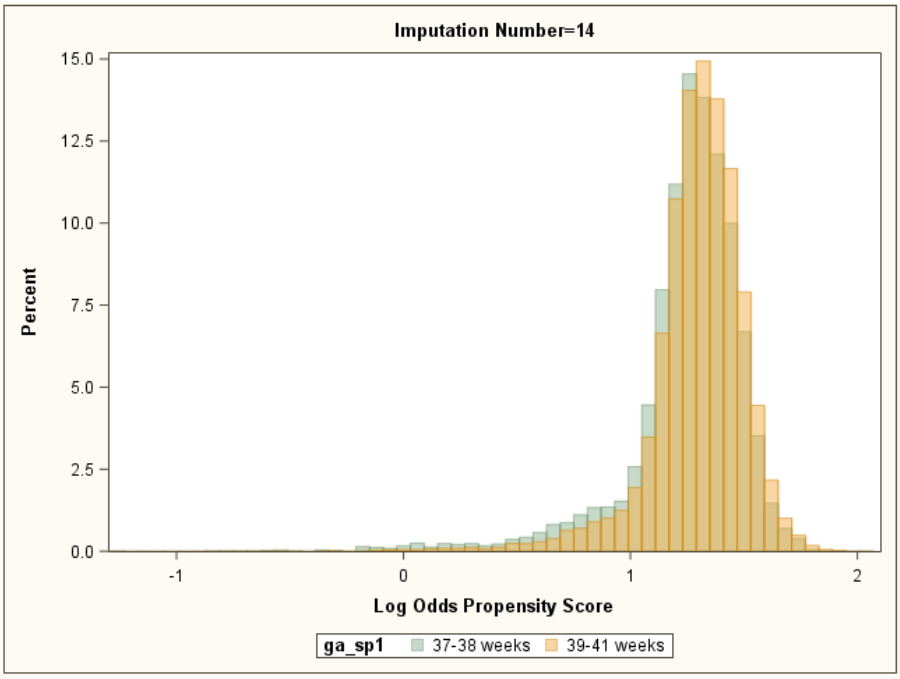

Imputation Number=9
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

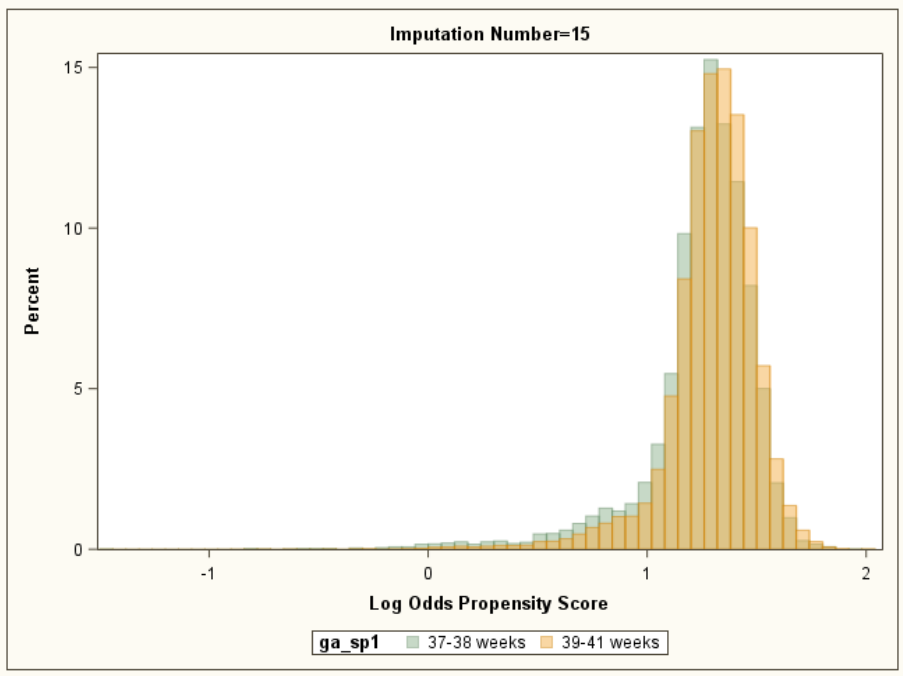

Imputation Number=10
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

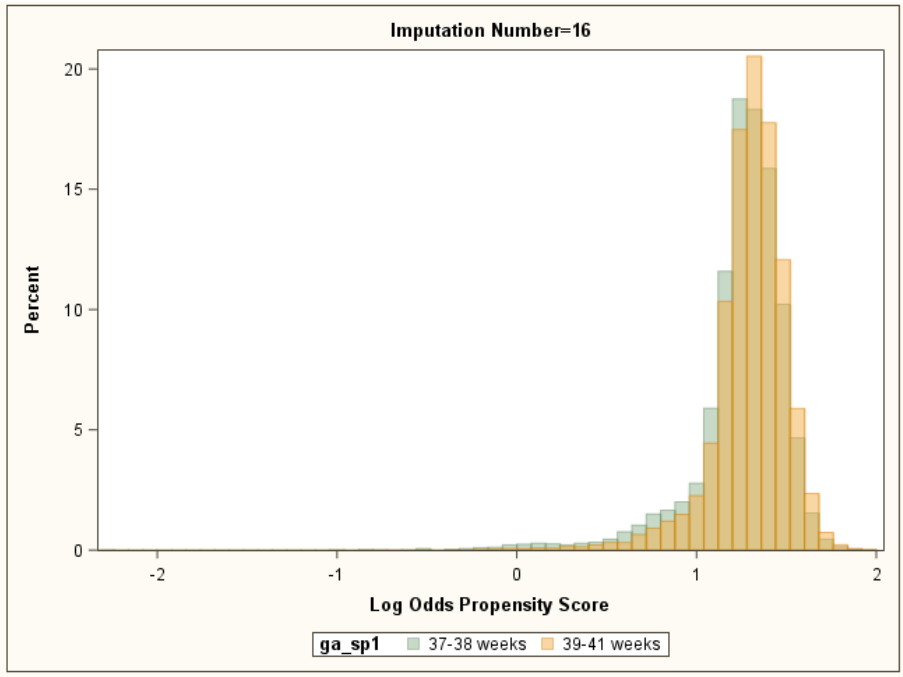

Imputation Number=11
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

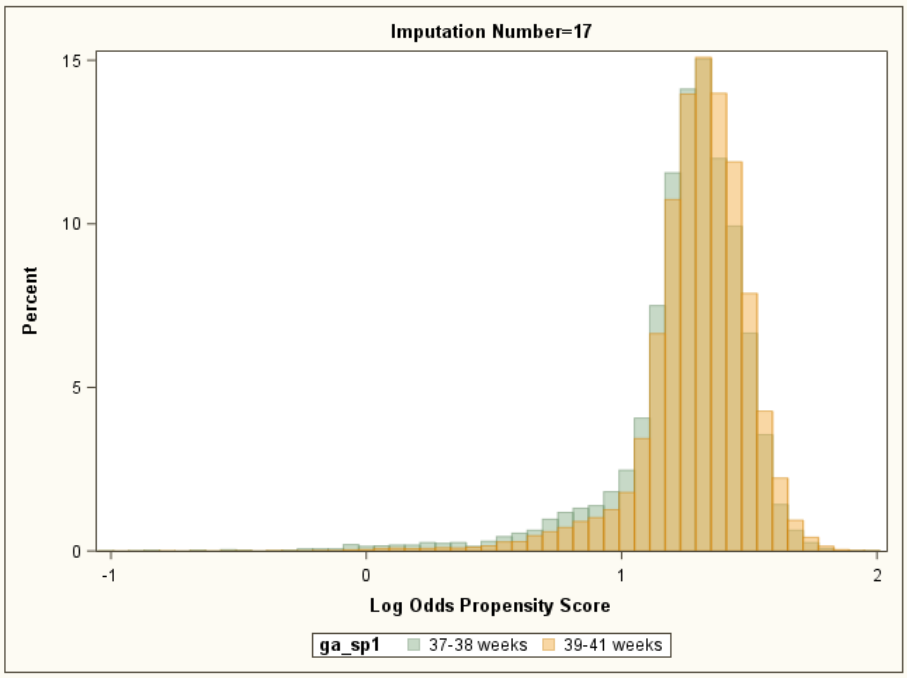

Imputation Number=12
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

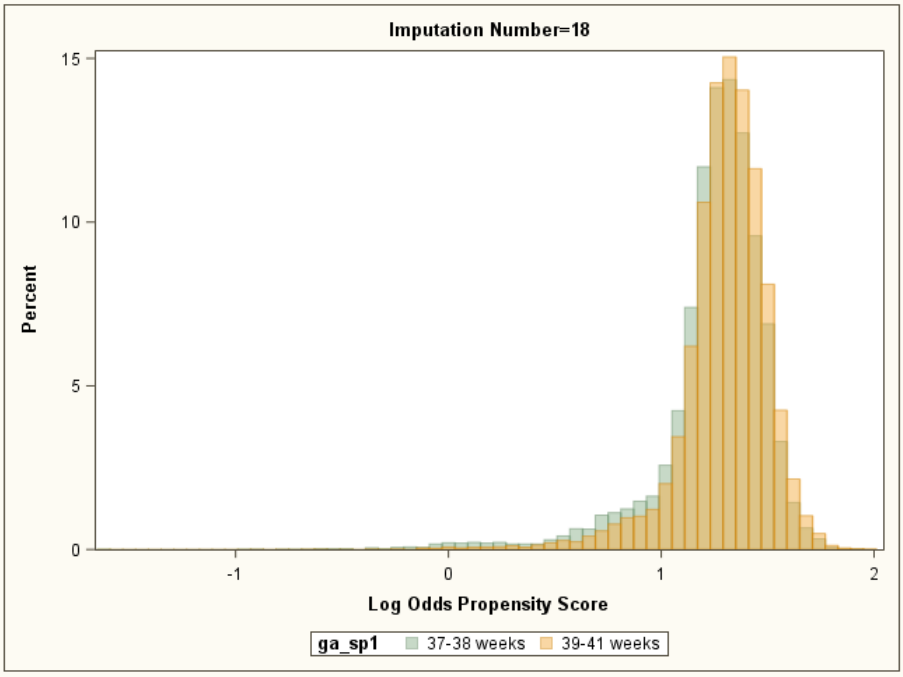

Imputation Number=13
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

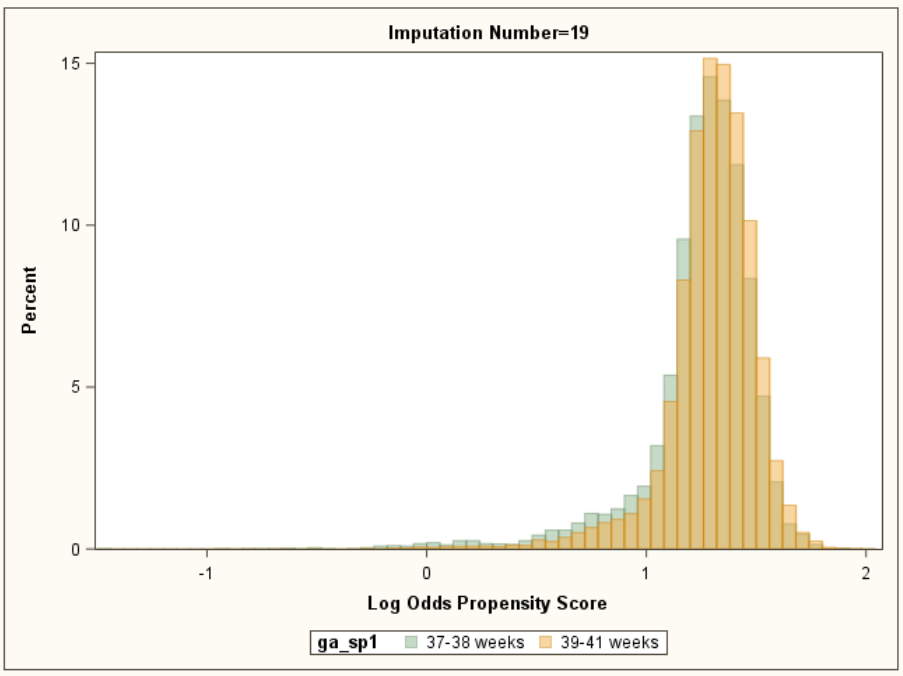

Imputation Number=14
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

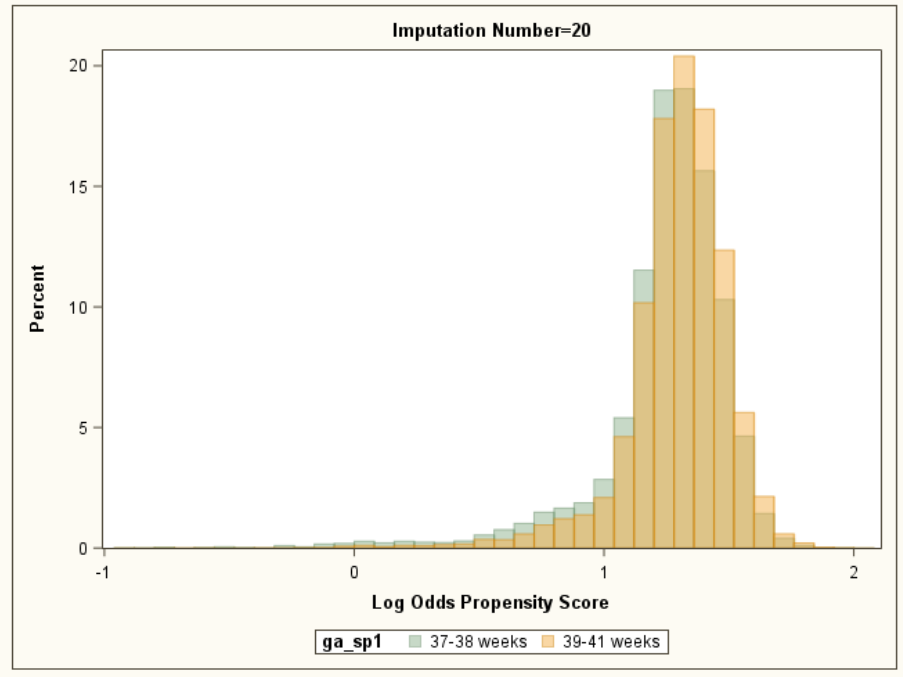

Imputation Number=15
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

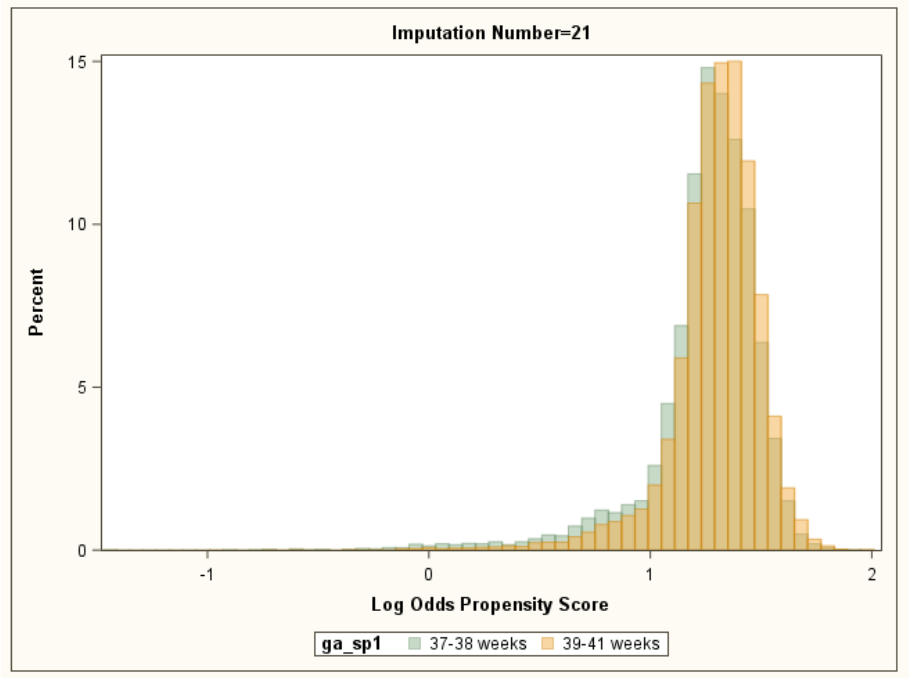

Imputation Number=16
Percent
20
15
10
5
0
-2
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

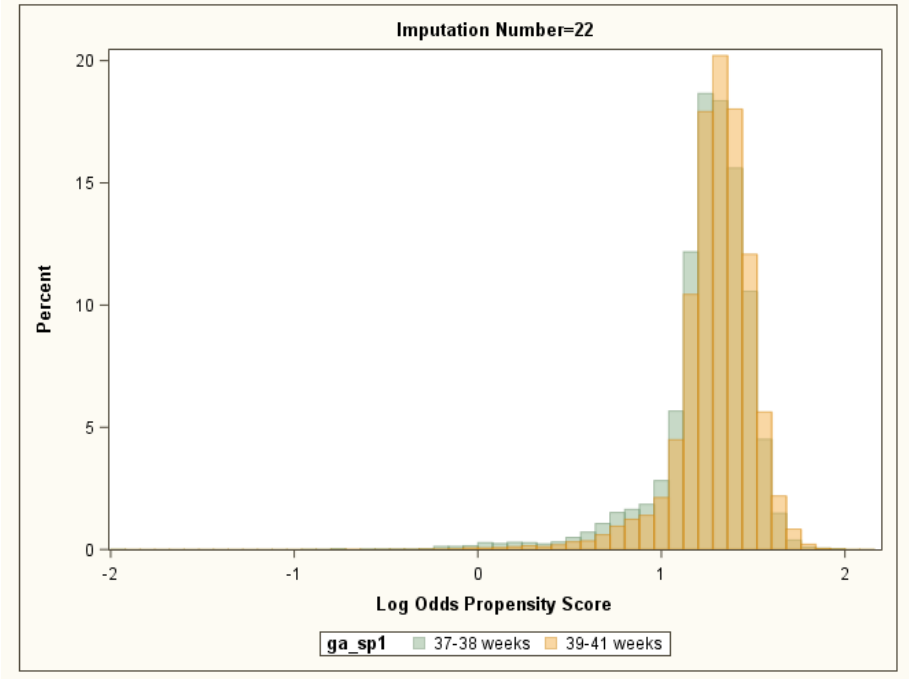

Imputation Number=17
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

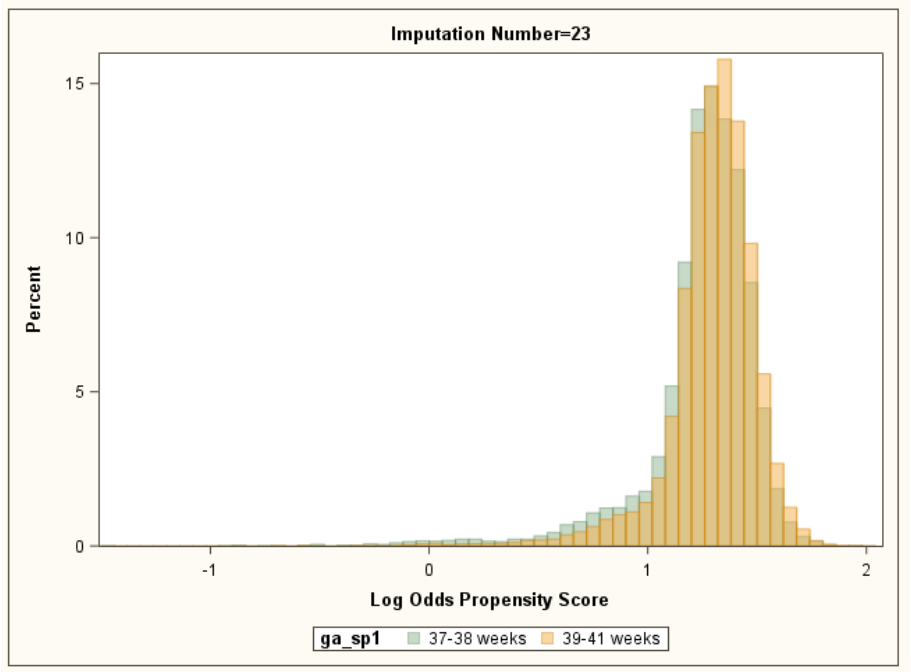

Imputation Number=18
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

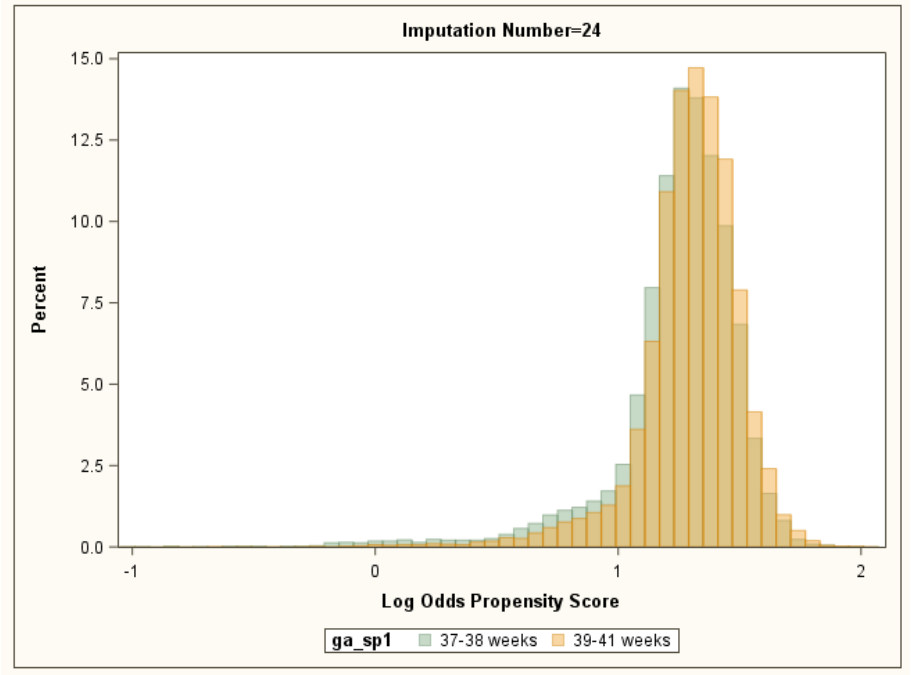

Imputation Number=19
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

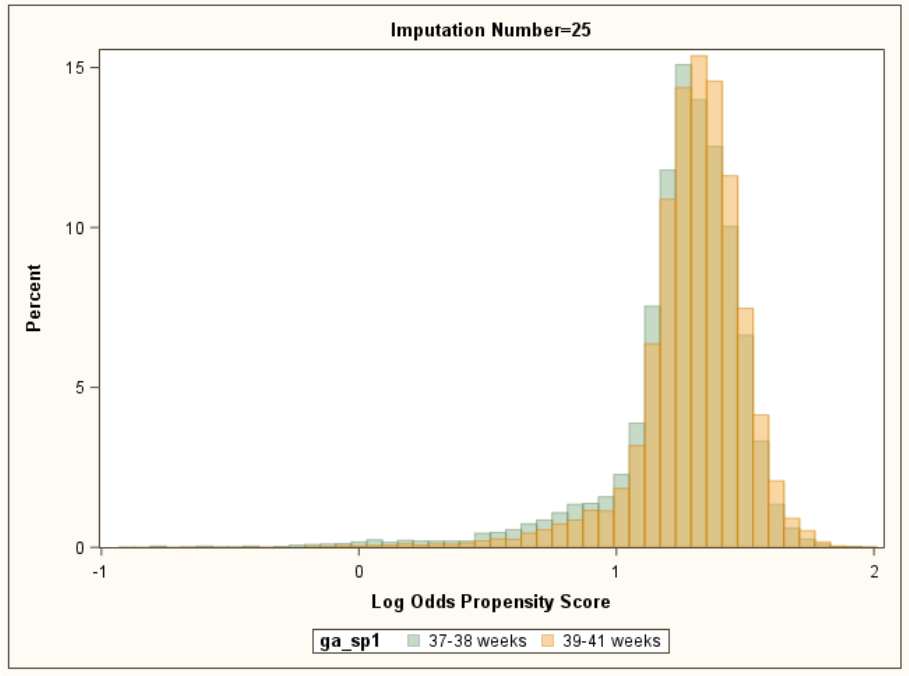

Imputation Number=20
Percent
20
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks


Imputation Number=21
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks


Imputation Number=22
Percent
20
15
10
5
0
-2
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks


Imputation Number=23
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks


Imputation Number=24
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks


Imputation Number=25
Percent
15
10
5
0
-1
0
1
2
Log Odds Propensity Score
ga_sp1
37-38 weeks
39-41 weeks

APPENDIX D. CORRELATIONS

Table D1. Correlations across 8 imputed datasets used in ELA z-score

Pair	r	Standard Error
BirthYear RaceNat	-0.001	0.006
BirthYear apgar5	-0.039	0.006
BirthYear lunch	-0.157	0.006
BirthYear max_testv_lvl	-0.098	0.006
BirthYear mth_del1	-0.005	0.006
BirthYear nicu	-0.208	0.009
BirthYear rnk	-0.449	0.014
BirthYear sga10	-0.058	0.006
apgar5 RaceNat	0.035	0.006
apgar5 lunch	-0.022	0.006
apgar5 max_testv_lvl	-0.013	0.006
apgar5 nicu	-0.167	0.007
apgar5 rnk	0.034	0.006
apgar5 sga10	-0.018	0.006
birthorder BirthYear	0.721	0.006
birthorder RaceNat	-0.012	0.006
birthorder apgar5	-0.021	0.006
birthorder lunch	-0.078	0.006
birthorder max_testv_lvl	-0.048	0.006
birthorder mth_del1	-0.033	0.006
birthorder nicu	-0.145	0.007
birthorder rnk	-0.467	0.015
birthorder sga10	-0.054	0.006
ga_sp1 BirthYear	-0.038	0.006
ga_sp1 RaceNat	0.016	0.006
ga_sp1 apgar5	0.017	0.006
ga_sp1 birthorder	-0.043	0.006
ga_sp1 lunch	-0.012	0.006
ga_sp1 max_testv_lvl	-0.007	0.006
ga_sp1 mth_del1	-0.010	0.006
ga_sp1 nicu	-0.033	0.006
ga_sp1 rnk	0.078	0.006
ga_sp1 sex1	0.016	0.006
ga_sp1 sga10	-0.010	0.006
lunch RaceNat	-0.027	0.006
lunch rnk	0.038	0.008
max_testv_lvl RaceNat	-0.101	0.006
max_testv_lvl lunch	0.110	0.007

Pair	r	Standard Error
max_testv_lvl rnk	0.036	0.007
max_testv_lvl sga10	0.026	0.006
mth_del1 RaceNat	-0.006	0.006
mth_del1 apgar5	-0.088	0.006
mth_del1 lunch	-0.058	0.006
mth_del1 max_testv_lvl	-0.034	0.007
mth_del1 nicu	0.070	0.007
mth_del1 rnk	0.001	0.006
mth_del1 sga10	-0.016	0.006
nicu RaceNat	-0.038	0.007
nicu lunch	0.073	0.007
nicu max_testv_lvl	0.055	0.007
nicu rnk	0.051	0.012
nicu sga10	0.068	0.007
rnk RaceNat	-0.018	0.008
sex1 BirthYear	0.009	0.006
sex1 RaceNat	-0.001	0.006
sex1 apgar5	0.006	0.006
sex1 birthorder	0.005	0.006
sex1 lunch	0.008	0.006
sex1 max_testv_lvl	-0.014	0.006
sex1 mth_del1	-0.028	0.006
sex1 nicu	-0.011	0.006
sex1 rnk	-0.011	0.006
sex1 sga10	-0.008	0.006
sga10 RaceNat	-0.011	0.006
sga10 lunch	0.032	0.006
sga10 rnk	-0.011	0.007

Table D2. Correlations across 12 imputed datasets used in Math z-score

Pair	r	Standard Error
BirthYear RaceNat	-0.001	0.006
BirthYear apgar5	-0.039	0.006
BirthYear lunch	-0.157	0.006
BirthYear max_testv_lvl	-0.099	0.006
BirthYear mth_del1	-0.005	0.006
BirthYear nicu	-0.206	0.010
BirthYear rnk	-0.443	0.012
BirthYear sga10	-0.058	0.006
apgar5 RaceNat	0.036	0.006
apgar5 lunch	-0.022	0.006
apgar5 max_testv_lvl	-0.011	0.006
apgar5 nicu	-0.168	0.006
apgar5 rnk	0.033	0.006
apgar5 sga10	-0.018	0.006
birthorder BirthYear	0.721	0.006
birthorder RaceNat	-0.012	0.006
birthorder apgar5	-0.021	0.006
birthorder lunch	-0.078	0.006
birthorder max_testv_lvl	-0.049	0.007
birthorder mth_del1	-0.033	0.006
birthorder nicu	-0.145	0.008
birthorder rnk	-0.461	0.012
birthorder sga10	-0.054	0.006
ga_sp1 BirthYear	-0.038	0.006
ga_sp1 RaceNat	0.016	0.006
ga_sp1 apgar5	0.017	0.006
ga_sp1 birthorder	-0.043	0.006
ga_sp1 lunch	-0.012	0.006
ga_sp1 max_testv_lvl	-0.006	0.006
ga_sp1 mth_del1	-0.010	0.006
ga_sp1 nicu	-0.032	0.007
ga_sp1 rnk	0.077	0.006
ga_sp1 sex1	0.016	0.006
ga_sp1 sga10	-0.010	0.006
lunch RaceNat	-0.027	0.006
lunch rnk	0.042	0.008
max_testv_lvl RaceNat	-0.102	0.006
max_testv_lvl lunch	0.111	0.006

Pair	r	Standard Error
max_testv_lvl rnk	0.035	0.006
max_testv_lvl sga10	0.026	0.006
mth_del1 RaceNat	-0.007	0.006
mth_del1 apgar5	-0.089	0.006
mth_del1 lunch	-0.058	0.006
mth_del1 max_testv_lvl	-0.035	0.007
mth_del1 nicu	0.068	0.007
mth_del1 rnk	0.000	0.006
mth_del1 sga10	-0.016	0.006
nicu RaceNat	-0.038	0.008
nicu lunch	0.072	0.006
nicu max_testv_lvl	0.051	0.008
nicu rnk	0.050	0.008
nicu sga10	0.068	0.009
rnk RaceNat	-0.017	0.007
sex1 BirthYear	0.009	0.006
sex1 RaceNat	-0.001	0.006
sex1 apgar5	0.006	0.006
sex1 birthorder	0.005	0.006
sex1 lunch	0.008	0.006
sex1 max_testv_lvl	-0.013	0.006
sex1 mth_del1	-0.027	0.006
sex1 nicu	-0.012	0.007
sex1 rnk	-0.011	0.006
sex1 sga10	-0.008	0.006
sga10 RaceNat	-0.011	0.006
sga10 lunch	0.032	0.006
sga10 rnk	-0.012	0.006

APPENDIX E. MODEL SUPPLEMENT

Table E1. Infinite Likelihood Warnings for Imputed Datasets

Model #	Imputed Data sets with Infinite Likelihood warning for ELA Model	Imputed Data sets with Infinite Likelihood warning for Math Model
4-Race/Ethnicity-Nativity	5,7,13,14,15,22,25	2,5,6,16,17,22
5-Race/Ethnicity-Nativity+Interaction	4,8,9,11,18,19,21,22,25	1,3,5,7,18,20,23

APPENDIX F. SELECTED SAS CODE

Propensity Score Model

```
proc glimmix data=ForPSmain noclprint method=LaPlace ;
  class siblingid;
  model ga_sp1 (descending)= W_Mage B_MAge_bc W_Fage B_Fage_bc
W_mEmploy B_mEmploy W_PrimFinanc B_PrimFinanc W_TobUse_recoded
B_TobUse_recoded W_diabetes B_diabetes
W_pregHTN B_pregHTN W_maternalrisk B_maternalrisk W_rsk_HTN B_rsk_HTN
W_rsk_PrevPre B_rsk_PrevPre W_rsk_renal B_rsk_renal W_Drug_any
B_Drug_any W_cmp_Cpro B_cmp_Cpro
W_cmp_Fdis B_cmp_FDis W_cmp_abrup B_cmp_abrup W_cmp_previa B_cmp_previa
W_NoPrenatalCare B_NoPrenatalCare W_AlcUse B_AlcUse W_Meduc_recoded
B_Meduc W_Feduc B_Feduc W_Congen B_Congen W_ObeseOverWt B_ObeseOverWt
         /cl dist=binary link=logit  ddfm=BETWITHIN solution covB;
  random intercept/subject=siblingid type=VC;
BY _imputation_;
ods output ParameterEstimates=PSFullMod CovB=covBPS
CovParms=PSFullCovPar FitStatistics=PSFull_fitstats;
 covtest/WALD;
 output out=glimPSCf pred( blup ilink)=PredProb_full
pred(noblup ilink)=PredProb_PA_full ; id _XBETA_ siblingid ga_sp1
_imputation_ childid;
*nloptions tech=newrap; nloptions gconv=0 TECH=NRRIDG;
title 'Full PS Model - by imputation';
 ods html;
run;
title; run;

proc mianalyze parms=PSFullMod;
modeleffects Intercept W_Mage B_MAge_bc W_Fage B_Fage_bc W_mEmploy
B_mEmploy W_PrimFinanc B_PrimFinanc W_TobUse_recoded B_TobUse_recoded
W_diabetes B_diabetes
W_pregHTN B_pregHTN W_maternalrisk B_maternalrisk W_rsk_HTN B_rsk_HTN
W_rsk_PrevPre B_rsk_PrevPre W_rsk_renal B_rsk_renal W_Drug_any
B_Drug_any W_cmp_Cpro B_cmp_Cpro
W_cmp_Fdis B_cmp_FDis W_cmp_abrup B_cmp_abrup W_cmp_previa B_cmp_previa
W_NoPrenatalCare B_NoPrenatalCare W_AlcUse B_AlcUse W_Meduc_recoded
B_Meduc W_Feduc B_Feduc W_Congen B_Congen W_ObeseOverWt B_ObeseOverWt;
run;

proc rank data=UseGlimPScf groups=5 out=glimPSCfR ;
ranks rnk;
var _XBETA_;
by _imputation_ ;
run;
```

Hybrid Random Effects Models – ELA z-score

Null Model

```
/*MODEL 0- NULL MODEL*/
 proc mixed data=ELATestFinalMods  method=ML covtest noclprint
PLOTS(MAXPOINTS=NONE) ;
  class siblingid;
```

```
  model z_ela=/ cl ddfm=BETWITHIN s covb;
  random intercept/subject=siblingid type=VC;
by _imputation_;
ods output SolutionF = ELA0fix FitStatistics=Fit0ela
CovParms=CovPar0ela;
ods html;
title 'ELA Model 0 -Unconditional (NULL) Model - by _imputation_';
  run;
title; run;
```

Final Model

```
/*MODEL 4 - Final Model*/
proc mixed data=ELATestFinalMods method=ML covtest noclprint ;
  class siblingid RaceNat ;
  model z_ela = B_ga_sp W_ga_sp B_sex W_sex B_birthorder W_birthorder
B_mth_del W_mth_del
B_apgar5 W_apgar5  B_nicu W_nicu B_max_testv_lvl W_max_testv_lvl
B_sga10 W_sga10 B_lunch W_lunch B_PredProb_full W_PredProb_full
RaceNat
/cl ddfm=BETWITHIN s covb;
  random intercept /subject=siblingid type=VC;
 lsmeans RaceNat / diff cl adjust=tukey;
by _imputation_ ;
 ods output SolutionF = ELA4fix CovB = ELA4covb CovParms=CovPar4ela
FitStatistics=Fit4ela
 lsmeans=lsmeans_4ela diffs=diff4ela;
ods html;
title 'ELA Model FINAL-Race-Nativity combo variable -no random slope';
  run;
title; run;
```

Hybrid Random Effects Models – Math z-score

Null Model

```
/*MODEL 0- NULL MODEL*/
 proc mixed data=MathTestFinalMods  method=ML covtest noclprint ;
  class siblingid;
  model z_math=/ cl ddfm=BETWITHIN s covb;
  random intercept/subject=siblingid type=VC;
by _imputation_;
 ods output SolutionF = Math0fix  CovParms=CovPar0math
FitStatistics=Fit0math;
ods html;
title 'MATH Model 0 -Unconditional (NULL) Model - by _imputation_';
  run;
title; run;
```

Final Model

```
/*MODEL 4 - Final Model*/
```

```
proc mixed data=MathTestFinalMods method=ML covtest noclprint;
  class siblingid RaceNat ;
  model z_math = B_ga_sp W_ga_sp B_sex W_sex B_birthorder W_birthorder
B_mth_del W_mth_del
B_apgar5 W_apgar5  B_nicu W_nicu B_max_testv_lvl W_max_testv_lvl
B_sga10 W_sga10 B_lunch W_lunch B_PredProb_full
W_PredProb_full RaceNat
/cl ddfm=BETWITHIN s covb;
  random intercept /subject=siblingid type=VC;
  lsmeans RaceNat / diff cl adjust=tukey;
by _imputation_ ;
    ods output SolutionF = Math4fix CovParms=CovPar4math
FitStatistics=Fit4math lsmeans=lsmeans_4math diffs=diff4math ;
ods html;
title 'MATH Model FINAL -Race-Nativity combo variable -no random
slope';
  run;
title; run;
```

APPENDIX G. BIRTH WEIGHT ANALYSIS

Table G1. Fixed Effects Predictors of ELA z-score among Study Sample Births, using birth weight as measure of birth size(n=31,647)

	Parameter	Estimate	Standard Error	95% CI		
	Fixed Effects					
	Intercept	0.266	0.150	-0.029	0.561	
Gestational Age	Between	-0.017	0.019	-0.055	0.020	
	Within	-0.013	0.015	-0.043	0.016	
Sex	Between	0.246	0.017	0.214	0.279	***
	Within	0.170	0.012	0.146	0.194	***
Birth Order	Between	-0.221	0.028	-0.275	-0.166	***
	Within	-0.094	0.012	-0.119	-0.070	***
Delivery Method	Between	0.045	0.019	0.008	0.082	*
	Within	0.013	0.027	-0.040	0.066	
Apgar score, 5 minutes	Between	0.052	0.014	0.025	0.080	**
	Within	-0.006	0.011	-0.028	0.016	
NICU admission	Between	-0.145	0.043	-0.240	-0.051	**
	Within	-0.029	0.025	-0.078	0.021	
Highest venous Blood Lead Level (bll)	Between	-0.023	0.002	-0.027	-0.018	***
	Within	-0.010	0.002	-0.013	-0.006	***
Birth Weight‡	Between	0.042	0.008	0.027	0.057	***
	Within	0.023	0.009	0.005	0.040	**
Ever school lunch eligible	Between	-0.500	0.021	-0.542	-0.458	***
	Within	0.002	0.029	-0.055	0.060	
Log Odds Propensity Score Rank	Between	0.045	0.009	0.025	0.065	***
	Within	0.015	0.008	-0.001	0.030	
Race/Ethnicity, Nativity	black,NL-NonUS	-0.368	0.026	-0.418	-0.317	***
	black,NL-US	-0.663	0.023	-0.708	-0.617	***
	white,NL-NonUS	-0.057	0.031	-0.117	0.004	
	white,NL-US		*ref*			
	Latino,NonUS	-0.409	0.023	-0.454	-0.364	***
	Latino,US	-0.556	0.023	-0.601	-0.511	***
	Asian/PI,NonUS	0.184	0.025	0.136	0.233	***
	Asian/PI,US	0.115	0.096	-0.074	0.304	
	Other,NonUS	-0.128	0.069	-0.263	0.008	
	Other,US	0.241	0.207	-0.164	0.647	

*p<0.05; **p<0.01;***p<0.001

‡Measured per 500 gram change

Table G2. Fixed Effects Predictors of Math z-score among Study Sample Births, using birth weight as measure of birth size (n=31,647)

	Parameter	Estimate	Standard Error	95% CI		
	Fixed Effects					
	Intercept	0.209	0.147	-0.079	0.497	
Gestational Age	Between	-0.029	0.019	-0.066	0.008	
	Within	-0.005	0.015	-0.035	0.025	
Sex	Between	0.036	0.016	0.004	0.068	*
	Within	0.004	0.012	-0.020	0.028	
Birth Order	Between	-0.147	0.027	-0.200	-0.094	***
	Within	-0.049	0.012	-0.073	-0.025	***
Delivery Method	Between	0.019	0.018	-0.016	0.054	
	Within	0.007	0.026	-0.044	0.058	
Apgar score, 5 minutes	Between	0.040	0.014	0.012	0.067	**
	Within	0.004	0.011	-0.018	0.026	
NICU admission	Between	-0.128	0.032	-0.191	-0.065	***
	Within	-0.040	0.025	-0.089	0.008	
Highest venous Blood Lead Level (bll)	Between	-0.017	0.002	-0.022	-0.013	***
	Within	-0.007	0.002	-0.011	-0.003	**
Birth Weight‡	Between	0.048	0.008	0.033	0.063	***
	Within	0.028	0.009	0.010	0.045	**
Ever school lunch eligible	Between	-0.375	0.021	-0.416	-0.334	***
	Within	0.030	0.030	-0.028	0.088	
Log Odds Propensity Score Rank	Between	0.037	0.006	0.025	0.049	***
	Within	0.021	0.007	0.008	0.034	**
Race/Ethnicity, Nativity	black,NL-NonUS	-0.306	0.025	-0.355	-0.256	***
	black,NL-US	-0.624	0.023	-0.669	-0.580	***
	white,NL-NonUS	0.054	0.030	-0.005	0.113	
	white,NL-US		*ref*			
	Latino,NonUS	-0.255	0.022	-0.299	-0.212	***
	Latino,US	-0.480	0.023	-0.524	-0.436	***
	Asian/PI,NonUS	0.413	0.024	0.365	0.460	***
	Asian/PI,US	0.226	0.095	0.039	0.412	*
	Other,NonUS	-0.025	0.068	-0.158	0.109	
	Other,US	0.403	0.203	0.004	0.802	*

*p<0.05; **p<0.01;***p<0.001

‡Measured per 500 gram change

Table G3. Random Effects Predictors of ELA z-score among Study Sample Births using birth weight as a measure of birth size (n=31,647)

Parameter	Estimate	Standard Error	95% CI		
RandomEffects					
Intercept(between sibling group variability)	0.239	0.007	0.226	0.253	***
Residual(within sibling group variability)	0.578	0.006	0.565	0.590	***
Fit Statistics					
AIC	81995.64				
-2 Log L	81931.64				

*p<0.05; **p<0.01;***p<0.001

Table G4. Random Effects Predictors of Math z-score among Study Sample Births using birth weight as a measure of birth size (n=31,647)

Parameter	Estimate	Standard Error	95% CI		
RandomEffects					
Intercept(between sibling group variability)	0.217	0.007	0.204	0.230	***
Residual(within sibling group variability)	0.589	0.007	0.576	0.602	***
Fit Statistics					
AIC	81807.00				
-2 Log L	81743.00				

*p<0.05; **p<0.01;***p<0.001

REFERENCES

1. Mathews T, MacDorman MF. Infant Mortality Statistics from the 2009 Period Linked Birth/Infant Death Data Set. 2013;*NCHS Data Brief* 61(8).

2. Wiingreen R, Greisen G, Svensson J, et al. Low gestational age at birth and difficulties in school—A matter of "dose."*PloS One*. 2018;13(6):e0198482.

3. Searle AK, Smithers LG, Chittleborough CR, et al. Gestational age and school achievement: a population study. *Archives of Disease in Childhood-Fetal and Neonatal Edition*. 2017;102(5):F409–F416.

4. Abel K, Heuvelman H, Wicks S, et al. Gestational age at birth and academic performance: population-based cohort study. *International Journal of Epidemiology*. 2017;46(1):324–335.

5. Williams BL, Dunlop AL, Kramer M, et al. Perinatal Origins of First-Grade Academic Failure: Role of Prematurity and Maternal Factors. *Pediatrics*. 2013;131(4):693–700.

6. Nepomnyaschy L, Hegyi T, Ostfeld BM, et al. Developmental Outcomes of Late-Preterm Infants at 2 and 4 Years. *Maternal and child health journal*. 2012;16(8):1612–1624.

7. Talge NM, Holzman C, Wang J, et al. Late-preterm birth and its association with cognitive and socioemotional outcomes at 6 years of age. *Pediatrics*. 2010;126(6):1124–1131.

8. Morse SB, Zheng H, Tang Y, et al. Early school-age outcomes of late preterm infants. *Pediatrics*. 2009;123(4):e622–e629.

9. Mathiasen R, Hansen BM, Andersen A-MNN, et al. Gestational age and basic school achievements: a national follow-up study in Denmark. *Pediatrics*. 2010;126(6):e1553–e1561.

10. Lipkind HS, Slopen ME, Pfeiffer MR, et al. School-age outcomes of late preterm infants in New York City. *American Journal of Obstetrics and Gynecology*. 2012;206(3):222–e1.

11. Kinney HC. The near-term (late preterm) human brain and risk for periventricular leukomalacia: a review. In: *Seminars in Perinatology*. 2006:81–88.

12. Peterson BS, Vohr B, Staib LH, et al. Regional brain volume abnormalities and long-term cognitive outcome in preterm infants. *JAMA:*2000;284(15):1939–1947.

13. Andrews JS, Ben-Shachar M, Yeatman JD, et al. Reading performance correlates with white-matter properties in preterm and term children. *Developmental Medicine and Child Neurology*. 2010;52(6):e94–e100.

14. Chan E, Quigley MA. School performance at age 7 years in late preterm and early term birth: a cohort study. *Archives of Disease in Childhood-Fetal and Neonatal Edition*. 2014;99(6):F451–F457.

15. American College of Obstetricians , Gynecologists, others. Definition of term pregnancy. Committee opinion No. 579. *Obstet Gynecol*. 2013;122:1139–1140.

16. Engle WA, Kominiarek MA. Late preterm infants, early term infants, and timing of elective deliveries. *Clinics in Perinatology*. 2008;35(2):325–341.

17. Mathews T, MacDorman MF, Thoma ME. Infant mortality statistics from the 2013 period linked birth/infant death data set. 2015; *NCHS Data Brief*

18. Chan E, Leong P, Malouf R, et al. Long-term cognitive and school outcomes of late-preterm and early-term births: a systematic review. *Child: Care, Health and Development*. 2016;42(3):297–312.

19. 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.

20. Robinson M, Whitehouse AJ, Zubrick SR, et al. Delivery at 37 weeks' gestation is associated with a higher risk for child behavioural problems. *Australian and New Zealand Journal of Obstetrics and Gynaecology*. 2013;53(2):143–151.

21. 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. *Archives of Disease in Childhood-Fetal and Neonatal Edition*. 2012;97(3):F167–F173.

22. Lindstrom K, Winbladh B, Haglund B, et al. Preterm infants as young adults: a Swedish national cohort study. *Pediatrics*. 2007;120(1):70–77.

23. Kirkegaard I, Obel C, Hedegaard M, et al. 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(4):1600–1606.

24. 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.

25. Yang S, Platt RW, Kramer MS. Variation in child cognitive ability by week of gestation among healthy term births. *American Journal of Epidemiology*. 2010;171(4):399–406.

26. Noble KG, Fifer WP, Rauh VA, et al. Academic achievement varies with gestational age among children born at term. *Pediatrics*. 2012;130(2):e257–e264.

27. MacKay DF, Smith GC, Dobbie R, et al. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS Medicine*. 2010;7(6):1–10.

28. Lawlor DA, Clark H, Smith GD, et al. Intrauterine growth and intelligence within sibling pairs: findings from the Aberdeen children of the 1950s cohort. *Pediatrics*. 2006;117(5):e894–e902.

29. Gisselmann M, Koupil I, De Stavola BL. The combined influence of parental education and preterm birth on school performance. *Journal of Epidemiology and Community*

Health. 2011;65(9):764–769.

30. Martin JA, Hamilton BE, Osterman MJK. Births in the United States, 2018. *NCHS Data Brief*. 2019;(346):1–8.

31. Martin JA. Births: Final Data for 2017. *NCHS Data Brief*. 2018;67(8).

32. Basso O, Wilcox A. Mortality risk among preterm babies: Immaturity vs. underlying pathology. *Epidemiology (Cambridge, Mass.)*. 2010;21(4):521.

33. Basso O, Wilcox AJ. Might rare factors account for most of the mortality of preterm babies? *Epidemiology (Cambridge, Mass.)*. 2011;22(3):320.

34. Heuvelman H, Abel K, Wicks S, et al. Gestational age at birth and risk of intellectual disability without a common genetic cause. *European Journal of Epidemiology*. 2018;33(7):667–678.

35. Mann JR, Pan C, Rao GA, et al. Children Born to Diabetic Mothers May be More Likely to Have Intellectual Disability. *Maternal and Child Health Journal*. 2012;:1–5.

36. Ying G, Liu C. Statistical analysis of clustered data using SAS system. In: *Proceedings of the North East SAS Users Group (NESUG) Conference*. 2006:17–20.

37. Lu MC, Halfon N. Racial and ethnic disparities in birth outcomes: a life-course perspective. *Maternal and Child Health Journal*. 2003;7(1):13–30.

38. Barker DJ. The fetal and infant origins of adult disease. *BMJ: British Medical Journal*. 1990;301(6761):1111.

39. Strauss RS. Adult functional outcome of those born small for gestational age. *JAMA: the journal of the American Medical Association*. 2000;283(5):625–632.

40. Marmot MG, Stansfeld S, Patel C, et al. Health inequalities among British civil servants: the Whitehall II study. *The Lancet*. 1991;337(8754):1387–1393.

41. Geronimus AT, Hicken M, Keene D, et al. "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *Journal Information*. 2006;96(5).

42. Williams DR. Race, socioeconomic status, and health the added effects of racism and discrimination. *Annals of the New York Academy of Sciences*. 1999;896(1):173–188.

43. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research. *JAMA: the journal of the American Medical Association*. 2005;294(22):2879–2888.

44. Williams DR. Racial/ethnic variations in women's health: the social embeddedness of health. *American Journal Of Public Health*. 2002;92(4):588–597.

45. Nepomnyaschy L. Socioeconomic gradients in infant health across race and ethnicity. *Maternal And Child Health Journal*. 2009;13(6):720–731.

46. Howerton CL, Bale TL. Prenatal programing: at the intersection of maternal stress and immune activation. *Horm Behav*. 2012;62(3):237–42.

47. Dunkel Schetter C, Tanner L. Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice. *Curr Opin Psychiatry.* 2012;25(2):141–148.

48. Flier JS, Underhill LH, McEwen BS. Protective and damaging effects of stress mediators. *New England Journal of Medicine*. 1998;338(3):171–179.

49. Buss C, Entringer S, Swanson JM, et al. The Role of Stress in Brain Development: The Gestational Environment's Long-Term Effects on the Brain. In: *Cerebrum: The Dana Forum On Brain Science*. 2012

50. Sokolovic N, Selvam S, Srinivasan K, et al. Catch-up growth does not associate with cognitive development in Indian school-age children. *European Journal Of Clinical Nutrition*. 2014;68(1):14.

51. Bright MA, Knapp C, Hinojosa MS, et al. The comorbidity of physical, mental, and developmental conditions associated with childhood adversity: a population based study. *Maternal And Child Health Journal*. 2016;20(4):843–853.

52. Lupien SJ, McEwen BS, Gunnar MR, et al. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* 2009;10(6):434–45.

53. Kurdek LA, Sinclair RJ. Relation of eighth graders' family structure, gender, and family environment with academic performance and school behavior. *Journal of Educational Psychology*. 1988;80(1):90.

54. Bradley RH, Caldwell BM, Rock SL. Home environment and school performance: A ten-year follow-up and examination of three models of environmental action. *Child Development*. 1988;:852–867.

55. Vernez G, Zimmer R. Interpreting the effects of Title I supplemental educational services. 2007;

56. Pfeiffer M, Slopen M, Curry A, et al. Creation of a Linked Inter-agency Data Warehouse: the Longitudinal Study of Early Development. 2012;

57. of Health NYCD, Hygiene M. New York City Childhood Lead Poisoning Prevention Program, Annual Report 2003. 2004;

58. Guo G, Harris KM. The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*. 2000;37(4):431–447.

59. Gurka MJ, LoCasale-Crouch J, Blackman JA. Long-term cognition, achievement, socioemotional, and behavioral development of healthy late-preterm infants. *Archives of Pediatrics And Adolescent Medicine*. 2010;164(6):525.

60. Smithers LG, Searle AK, Chittleborough CR, et al. A whole-of-population study of term and post-term gestational age at birth and children's development. *BJOG*.

2015;122(10):1303–11.

61. Bilsteen JF, Taylor-Robinson D, Borch K, et al. Gestational age and socioeconomic achievements in young adulthood: a Danish population-based study. *JAMA Network Open*. 2018;1(8):e186085.

62. Dornbusch SM, Ritter PL, Leiderman PH, et al. The relation of parenting style to adolescent school performance. *Child Development*. 1987;:1244–1257.

63. Yang S, Lynch J, Susser ES, et al. Birth weight and cognitive ability in childhood among siblings and nonsiblings. *Pediatrics*. 2008;122(2):e350–e358.

64. Matte TD, Bresnahan M, Begg MD, et al. Influence of variation in birth weight within normal range and within sibships on IQ at age 7 years: cohort study. *BMJ*. 2001;323(7308):310–314.

65. Lawlor DA, Bor W, O'Callaghan MJ, et al. Intrauterine growth and intelligence within sibling pairs: findings from the Mater-University study of pregnancy and its outcomes. *Journal Of Epidemiology And Community Health*. 2005;59(4):279–282.

66. Eriksen W, Sundet JM, Tambs K. Birth weight standardized to gestational age and intelligence in young adulthood: a register-based birth cohort study of male siblings. *American Journal Of Epidemiology*. 2010;172(5):530–536.

67. Dobbing J, Sands J. Quantitative growth and development of human brain. *Archives of Disease in Childhood*. 1973;48(10):757–767.

68. Najman JM, Aird R, Bor W, et al. The generational transmission of socioeconomic inequalities in child cognitive development and emotional health. *Social Science and Medicine*. 2004;58(6):1147–1158.

69. Ananth CV, Schisterman EF. Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics. *American Journal Of Obstetrics And Gynecology*. 2017;217(2):167–175.

70. Bandoli G, Palmsten K, Flores KF, et al. Constructing causal diagrams for common perinatal outcomes: benefits, limitations and motivating examples with maternal antidepressant use in pregnancy. *Paediatric And Perinatal Epidemiology*. 2016;30(5):521–528.

71. Hutcheon JA, Moskosky S, Ananth CV, et al. Good practices for the design, analysis, and interpretation of observational studies on birth spacing and perinatal health outcomes. *Paediatric And Perinatal Epidemiology*. 2019;33(1):O15–O24.

72. Luque-Fernandez MA, Schomaker M, Redondo-Sanchez D, et al. Educational Note: Paradoxical collider effect in the analysis of non-communicable disease epidemiological data: a reproducible illustration and web application. *International Journal Of Epidemiology*. 2018;48(2):640–653.

73. New York City Office of Vital Statistics and, Epidemiology. Annual Summary of Vital Statistics, 1997. December,1997;

74. Bell A, Fairbrother M, Jones K. Fixed and random effects models: making an informed choice. *Quality and Quantity*. 2019;53(2):1051–1074.

75. Carlin JB, Gurrin LC, Sterne JA, et al. Regression models for twin studies: a critical review. *International Journal of Epidemiology*. 2005;34(5):1089–1099.

76. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Ed. Thousand Oaks (CA): Sage Publications; 2002. (Advanced Quantitative Techniques in the Social Sciences; vol 1).

77. Ruff R. Hierarchical Linear Modeling in Public Health Research. 2014[unpublished presentation]

78. Schempf, AH . Fixed Versus Random Effects Models for Multilevel and Longitudinal Data Analysis. 2012;(https://www.mchnavigator.org/trainings/1338-epi-2012-course/fixed-random.php?page_id=3746). (Accessed January, 2018)

79. Schunck R. Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*. 2013;13(1):65–76.

80. Schunck R, Perales F. Within-and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. *The Stata Journal*. 2017;17(1):89–115.

81. Fielding A, Goldstein H. Cross-classified and multiple membership structures in multilevel models: An introduction and review. 2006;(791).

82. Mann V, De Stavola BL, Leon DA. Separating within and between effects in family studies: an application to the study of blood pressure in children. *Statistics in Medicine*. 2004;23(17):2745–2756.

83. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*. 1998;24(4):323–355.

84. Berglund P, Heeringa SG. Multiple imputation of missing data using SAS. SAS Institute; 2014.

85. Donders ART, Van Der Heijden GJ, Stijnen T, et al. A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*. 2006;59(10):1087–1091.

86. Gelman A, Hill J. Missing-data imputation. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press; 2007:529–543.

87. Azur M, Constantine Frangakis, Stuart E. Disparities Among Children Served by the CMHS Children's Services Program: Overview of Multiple Imputation and Using

Multiply Imputed Data. Johns Hopkins School of Public Health;2008.

88. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*. 2017;9:157–166.

89. UCLA Group. Multiple Imputation in SAS Part 1. (https://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/). (Accessed December 20, 2018)

90. Grund S, Ludtke O, Robitzsch A. Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*. 2018;21(1):111–149.

91. Grund S, Lüdtke O, Robitzsch A. Multiple imputation of multilevel missing data: An introduction to the R package pan. *SAGE Open*. 2016;6(4):1–17.

92. Birrell CL. Handling Missing Categorical Data in a Multi-level Structure. 2017;

93. Enders CK, Keller BT, Levy R. A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*. 2018;23(2):1–19.

94. Lloyd JE, Obradovic J, Carpiano RM, et al. JMASM 32: Multiple imputation of missing multilevel, longitudinal data: A case when practical considerations trump best practices? *Journal of Modern Applied Statistical Methods*. 2013;12(1):261–275.

95. Hua He, He J. Overview of Propensity Score Methods. In: He, H and Wu, P and Chen, D.-G.D., ed. *Statistical Causal Inferences and Their Applications in Public Health Research*. Springer International Publishing; 2016:29–48.

96. Phuong NNL. Essays on Propensity Score Methods for Causal Inference in Observational Studies. 2018.[dissertation]

97. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.

98. Xiang Y, Wang S. An application of propensity score stratification using multilevel models: Do charter schools make a difference in student achievement and growth. *Portland: Northwest Evaluation Association*. 2013;

99. Stokes ME, Davis MESCS, Koch GG, et al. Categorical data analysis using the SAS system. SAS institute; 2003.

100. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *American Journal of Epidemiology*. 2006;163(12):1149–1156.

101. Rosenberg D, Rankin K. Propensity Scores. 2012;(https://www.mchnavigator.org/trainings/1338-epi-2012-course/fixed-random.php?page_id=3744)

102. Schabenberger O. Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models. 2005;

103. Lawlor DA, Lichtenstein P, Fraser A, et al. Does maternal weight gain in pregnancy have long-term effects on offspring adiposity? A sibling study in a prospective cohort of 146,894 men from 136,050 families. *The American Journal of Clinical Nutrition*. 2011;94(1):142–148.

104. A. BP. Using SAS for Multiple Imputation and Analysis of Longitudinal Data. 2018;(1738).

105. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*. 2006;101(475):901–910.

106. Sulloway FJ. Birth order and intelligence. *Science*. 2007;316(5832):1711–1712.

107. Miller RT, Murnane RJ, Willett JB. Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis*. 2008;30(2):181–200.

108. De Bie HM, Oostrom KJ, Boersma M, et al. Global and regional differences in brain anatomy of young children born small for gestational age. *PloS One*. 2011;6(9):e24116.

109. Vollmer B, Edmonds CJ. School Age Neurological and Cognitive Outcomes of Fetal Growth Retardation or Small for Gestational Age Birth Weight. *Frontiers in Endocrinology*. 2019;10(186):1–17.

110. Miller SL, Huppi PS, Mallard C. The consequences of fetal growth restriction on brain structure and neurodevelopmental outcome. *The Journal of Physiology*. 2016;594(4):807–823.

111. Hill MA. Embryology Neural System - Fetal. (https://embryology.med.unsw.edu.au/embryology/index.php/Neural_System_-_Fetal). (Accessed December 21, 2019)

112. Hill MA. Embryology Neural - Cerebrum Development. (https://embryology.med.unsw.edu.au/embryology/index.php/Neural_-_Cerebrum_Development). (Accessed December 21, 2019)

113. Breining S, Doyle J, Figlio D, et al. Birth order and delinquency: Evidence from Denmark and Florida. 2017;

114. Evens A, Hryhorczuk D, Lanphear BP, et al. The impact of low-level lead toxicity on school performance among children in the Chicago Public Schools: a population-based retrospective cohort study. *Environmental Health*. 2015;14(21):1–9.

115. Reardon SF, Zárate RC. The Relationship Between Test Item Format and Gender Achievement Gaps on Math and ELA Tests in 4th and 8th Grade. *Educational Researcher,*. 2018;47(5):284–294.

116. Rohrer JM, Egloff B, Schmukle SC. Examining the effects of birth order on personality. *Proc. Natl. Acad. Sci. U.S.A.* 2015;112(46):14224–9.

117. Andreias L, Borawski E, Schluchter M, et al. Neighborhood influences on the academic achievement of extremely low birth weight children. *Journal of Pediatric Psychology*. 2009;35(3):275–283.

118. Yang S, Kestens Y, Dahhou M, et al. Neighborhood deprivation and maternal psychological distress during pregnancy: A multilevel analysis. *Maternal and Child Health Journal*. 2015;19(5):1142–1151.

119. Kelly AB, O'Flaherty M, Connor JP, et al. The influence of parents, siblings and peers on pre- and early-teen smoking: a multilevel model. *Drug Alcohol Rev*. 2011;30(4):381–7.

120. of Obstetricians AC, Gynecologists, others. Nonmedically indicated early-term deliveries. Committee Opinion No. 561. *Obstet Gynecol*. 2013;121(4):911–915.

121. American Community Survey, 2008 3-year estimates. (https://www.census.gov/programs-surveys/acs/data.html). (Accessed November, 2019)

122. DiGiuseppe DL, Aron DC, Ranbom L, et al. Reliability of birth certificate data: a multi-hospital comparison to medical records information. *Maternal and Child Health Journal*. 2002;6(3):169–179.

123. Martin JA, Wilson EC, Osterman MJ, et al. Assessing the Quality of Medical and Health Data From the 2003 Birth Certificate Revision: Results From Two States. 2013;(vol 62 no 2).

124. Reichman NE, Hade EM. Validation of birth certificate data: a study of women in New Jersey's HealthStart program. *Annals of Epidemiology*. 2001;11(3):186–193.

125. Frisell T, Oberg S, Kuja-Halkola R, et al. Sibling Comparison Designs: Bias From Non-Shared Confounders and Measurement Error. *Epidemiology*. 2012;23(5):713–720.

126. Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*. 2019;56(6):774–786.

127. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. *Environmental Health Perspectives*. 2011;119(6):878–885.

128. Mitro SD, Johnson T, Zota AR. Cumulative chemical exposures during pregnancy and early development. *Current Environmental Health Reports*. 2015;2(4):367–378.

129. Bell BA, Owens CM, Ferron JM, et al. Parsimony vs. complexity: a comparison of two-level, three-level, and cross-classified models using add health and AHAA data. *SESUG Proceedings (c) SESUG, Inc (http://www. sesug. org) Paper PO091 [cited 2013 May 13]. Available from: http://analytics. ncsu. edu/sesug/2008/PO-091. pdf*. 2013;

130. Dunn E, Miliren C, Evans C, et al. Using cross-Classified Multilevel Models to Disentangle the Relative Influence of Schools and Neighborhoods on Adolescents' Depressive Symptoms. *International Journal of Epidemiology*. 2015;44(suppl_1):i49–i49.

131. Leckie G. Module 12: cross-classified multilevel models . 2013;UK Bristol [online manual] (accessed, December 2018)

132. Schunck R. Cluster size and aggregated level 2 variables in multilevel models. A cautionary note. *Methods, Data, Analyses*. 2016;10(1):97–108.

133. Livingston G, Parker K, Rohal M. Childlessness falls, family size grows among highly educated women. *Pew Research Center*. 2015;

134. Parikh LI, Reddy UM, Männistö T, et al. Neonatal outcomes in early term birth. *American Journal of Obstetrics and Gynecology*. 2014;211(3):265–e1.

135. Tita AT, Landon MB, Spong CY, et al. Timing of elective repeat cesarean delivery at term and neonatal outcomes. *New England Journal of Medicine*. 2009;360(2):111–120.

136. Bell K, Bloor K, Hewitt C. How Do Undiagnosed Symptoms of Maternal Psychological Distress During the Postnatal Period Affect Child Developmental Outcomes? *Maternal and Child Health Journal*. 2019;23:1187–1195.