INTERVENTIONAL PHYSICAL AND OCCUPATIONAL THERAPY SERVICES AND MOTOR COORDINATION AMONG LOW BIRTH WEIGHT INFANTS

Stephanie Elaine Watkins

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in the Department of Epidemiology, Gillings School of Public Health.

Chapel Hill
2012

Approved by:

Julie Daniels PhD

Michele Jonsson-Funk PhD

Alan Brookhart PhD

Michael O'Shea MD, MPH

Steven A. Rosenberg PhD

Abstract

STEPHANIE ELAINE WATKINS: Interventional Physical and Occupational Therapy Services and Motor Coordination among Low Birth Weight Infants
(Under the direction of Julie Daniels)

Introduction: Children born very low birth weight (VLBW) have an increased risk of impaired preschool motor coordination, which may have negative effects on the child's mental and physical health. Physical and occupational therapy services are suggested to attenuate the negative effects of poor preschool coordination. We estimated the effect of physical and occupational therapy services delivered in early childhood on preschool motor coordination among VLBW children. To control for confounding, we implemented propensity score (PS) methods estimated using traditional logistic regression (LR) and tree based methods. Methods: Using the Early Childhood Longitudinal Study Birth Cohort (ECLS-B) we estimated the effect of therapy on: skipping eight consecutive steps, hopping five times, standing on one leg for ten seconds, walking backwards six steps on a line, jumping distance, and change in jumping distance from preschool to kindergarten. We estimated the PS using random forest classification, bagging, and a single tree using the R statistical program and with LR in SAS 9.2. Using linear regression, we modeled the estimated effect of therapy on the distance that the child jumped. We weighted the adjusted models using inverse probability of treatment weights estimated from all four methods. We modeled all other end points as stated using LR. Results: Approximately 500 children were VLBW. RF and Bagging produced the best covariate balance between treatment groups (MSD 0.07, 0.03). The single classification tree produced the worst covariate balance (MDS 0.18). When estimating the PS with RF, treated VLBW children were 2.39 times as likely to successfully skipping eight steps (OR: 2.39, 95% CI: 0.75, 7.51) compared to the untreated group. Treated children jumped an additional 1.79 inches (95% CI: -2.21-5.79) further and were also 52% (OR: 1.52, 95% CI: 0.51, 4.54) more likely to successfully complete the backwards walking task. There was little effect of therapy on other endpoints. Effect estimates were similar among models weighted with RF, bagging, and LR. Conclusion: Providing therapy to VLBW children, may improve the child's school age motor coordination. RF is a useful method to improve covariate balance when estimating the PS and to potentially reduce bias in observational studies.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figures

Chapter 1

BACKGROUND AND SIGNIFICANCE

Statement of Specific Aims

Since the 1980's, rates of infants born with low birth weight (LBW) have increased in the United States. In 2010, approximately 8.2% of all births were LBW and 1.5% of all births were very low birth weight (VLBW).[1]

Approximately 10% of children who are low birth weight exhibit severe neurological impairments including abnormalities in tone, transitional movement, and persistence of primitive reflexes.[2,3] However many low birth weight children only experience minor neurological impairments.

Children born VLBW without notable neurological deficits often show an initial delay in foundational motor skills. However, these children often catch up to children of normal birth weight during the first few years of life. Although they typically attain foundational motor milestones, at school age these children are challenged to learn new motor tasks involving balance and coordination. They may appear "clumsy" or "awkward" and have difficulty with daily activities and classroom skills such as tying shoes or participating in physical education.[4] Poor motor coordination in childhood has a negative effect on the child's mental and physical health with persists into adolescence.[5,6-9]

In the absence of a neurological or medical diagnosis, these symptoms are described as developmental coordination disorder (DCD).[10] The prevalence of DCD among VLBW/very preterm infants has been reported to be as high as 72%. Children born with VLBW are six times as likely to develop DCD than children born with normal birth weight.[10]

To attenuate the potential negative sequelae of poor childhood motor coordination, early intervention (EI) by physical and occupational therapists is recommended.[10] EI is a federal program providing interventional services to infants and toddlers to improve outcomes for children with developmental disabilities. Specifically, physical and occupational therapists often treat low birth weight children to improve function and to minimize morbidity during childhood.

In the published literature, few studies have examined the efficacy of interventional physical and occupational therapy services on school age motor skills of low birth weight children. Previous research comprises a heterogeneous group of studies where small groups of preterm and low birth weight infants are typically randomized to neurodevelopmental treatment or typical care. The majority of the studies evaluate interventional effects within the first twelve months of life with variability in both intensity of treatment and length of follow up.

In a population of VLBW infants without neurological involvement, two small randomized studies assessed the effect of occupational therapy and physical therapy during the first year of life on motor ability in childhood.[11,12] Neither trial found a statistically significant difference in standardized motor scores between treatment groups. However, among children born of normal birth weight, interventions promoting motor development appear to improve children's locomotor ability.[13]

Little is known regarding the efficacy of physical and occupational therapy on childhood motor coordination. Observational data are available to evaluate this relationship, yet research analyzing the effect of a treatment on an outcome in non-randomized studies is complicated by exposure group differences on measured and unmeasured characteristics associated with the outcome of interest.

Propensity scores (the predicted probability of treatment given a set of measured covariates) are a commonly used method to control for confounding when estimating the average treatment effect in observational studies, yet there are few guidelines in the literature regarding how to estimate the propensity score.[14] Logistic regression is frequently used, yet the model must conform to the assumption of linearity between ordinal and continuous covariates and the logit of the dependent variable. Furthermore the joint effect between independent variables in the model must be considered, as well as the functional form of covariates or interaction terms.[15] Violations can result in misspecification of the propensity score model and the resulting effect estimate may be biased.[16]

Tree based methods, including Bagging and Random Forest classification (RFC), are non-parametric methods derived from learning based algorithms which offer robust alternative strategies for generating predicted probabilities of treatment.[17,18] Yet, these methods are underutilized in the literature.

Using the Early Childhood Longitudinal Study Birth Cohort (ECLS-B) (N≈1,150), this study will estimate the association between receipt of interventional physical and occupational therapy services and motor coordination during preschool and school age developmental periods in a population of children born VLBW. Furthermore, we will consider the use of novel methods, propensity score estimation using tree based methods, to control for confounders in these data.

Specific Aim 1: In a population of very low birth weight children, we will estimate the effect of early childhood physical and occupational therapy services on preschool age motor coordination.

Specific AIM 2: We will illustrate two novel methods, random forest classification and bagging, to estimate the predicted probability of receiving early childhood physical or occupational therapy. We will compare these methods with other tree based methods as well as logistic regression with regard to covariate balance, bias, and precision of the estimated effect of therapy on preschool motor performance.

Overview

Since the 1980's, rates of infants born with low birth weight (LBW) have increased in the United States. In 2010, approximately 8.2% of all births were LBW and 1.5% of all births were very low birth weight (VLBW).[1] Approximately 67% of low birth weight babies are also born preterm.[19]

Low birth weight children are at risk for long term morbidity and developmental disability.[20,21] Especially among LBW infants who are also born preterm, these infants may experience major disturbances during a period of rapid brain growth which may result in abnormalities in tone and movement patterns.[22] These babies may have a poor ability to assume flexion and frequently maintain patterns of extension. These abnormal movement patterns often lead to delays in unsupported sitting and trunk rotation which in turn affects fine motors skills, behavior, and cognition.[23]

A small percentage of VLBW children (≈10%) suffer from these severe neurological impairments that affect posture and movement. Yet, a large proportion experience only minor motor difficulties with complex movement later in development.[2,3] These impairments in motor coordination may have negative effects on the child's self-esteem and level of physical activity which may persist into adolescence.[5-7]

To minimize dysfunction and disability among children born LBW, early intervention programs are often implemented. Early Intervention is a federal program that delivers services to infants and toddlers through three years of age.[24] Physical therapy, occupational therapy, and speech therapy are services directly delivered to the child and the family. These services are frequently referred to as “interventional therapy services”.

I. Motor Development in Low Birth Weight and Preterm Infants

Over the last twenty years, a large body of research explored impairments in motor development among low birth weight and preterm children. Since a large percentage of low birth weight children are also preterm, we consider the implications of both factors on motor development.[19] Impairments in motor ability not only affect a child’s ability to move about their environment but also handwriting, behavior, and cognitive performance.[25-28] Movement is a child’s connectivity to the world. It is through exploration of their environment that children learn.

The degree of impairment among children born low birth weight can be quite variable. Some children in this population exhibit severe motor delays with little independent movement, where other children experience milder delays in motor coordination.

Approximately 10% of low birth weight preterm children develop cerebral palsy; a disorder of posture and movement.[2,29-31] These children have abnormalities in tone and transitional movement with persistence of primitive reflexes. Abnormalities in motor control lead to delays in motor milestone attainment and functional ability. Typically these motor abnormalities are associated with perinatal brain hypoxia, ischemia, infection, intraventricular hemorrhage, and periventricular leukomalacia.[3]

However, not all low birth weight preterm children develop such severe impairments in motor function. Many children who are born low birth weight show only minor or no obvious neurological impairments.

Children that are born low birth weight perform more poorly on assessments of motor performance compared to children of normal birth weight. Furthermore, their motor performance appears to decline with increasing prematurity.[32]

Between 1992 and 2009, over 24 studies examined motor development among very low birth weight and very preterm children. When compared to a normative sample, VLBW and very preterm children scored lower on the psychomotor developmental index (PDI) of the Bayley II. These children were on average 0.88 standard deviations behind their typically developing peers (95% CI -0.96 to -0.80).[33] When researchers examined the PDI score for children with adverse perinatal complications, the effect size decreased further (0.51 SD).[33]

VLBW and very preterm children also demonstrate difficulty with higher level motor skills of balance and coordination. These soft signs of motor impairments are often seen in school age children.[34] Researchers commonly evaluate level of impairment using one of the most recognized instruments of motor performance; The Movement Assessment Battery for Children. Compared to their full term peers, very low birth weight or very preterm children consistently have a higher overall impairment score. Moreover, they scored lower on the following subscales: balls skills; balance skills; and manual dexterity.[33,35] Children exhibited the greatest deficit in balance skills. Compared to a normative sample, children in this population also have lower motor proficiency scores in running speed, agility, coordination, strength, and dexterity.[33]

Although initially delayed, very preterm and very low birth weight children exhibited a catch up effect in early childhood. Yet, they exhibited a deficit in more complex motor tasks, as measured by the Movement Assessment Battery for Children, as they moved into school age and adolescence.[33] The divergence of skills began to appear at age five as children entered into elementary school.

Children with delays and difficulty in motor coordination may be described as having Developmental Coordination Disorder (DCD). This condition is described as “marked impairment in the development of motor coordination” among children without a known neurological or medical condition which describes their incoordination.[36] Symptoms of the condition are often first noticed during preschool when the child first attempts to learn movement requiring balance and coordination. Prior to preschool, the majority of these children were meeting normal developmental milestones.[4] Children with this condition have difficulty with new motor tasks and execution of coordinated movement. These children often avoid activities which require bilateral balance and coordination. Children with DCD are at risk for low academic performance, low self-esteem, and limited participation in physical activity.[37] Very low birth weight children are 6 times as likely to have DCD when compared to their normal birth weight peers.[10]

II. Early intervention

Early intervention physical and occupational therapy services are recommended to facilitate motor control among low birth weight children who are at risk for delays in gross motor skills. Early intervention describes a group of services and programs, provided to children with developmental delays, to improve their functional ability. In 1986, under the Individual with Disabilities Education Act, the United States Congress passed Public Law 105-17 mandating the provision of infant and toddler early intervention services. These early intervention services are called Part C.[24]

Part C intervention programs are multidisciplinary statewide programs which operate within the guidelines set by the federal government. Specific eligibility criteria are set by each state. Part C programs provide services for children from birth to age three. Children generally qualify if they have documented impairments in one or more of the following developmental areas: *motor, cognitive, adaptive, communicative, social, or emotional.*

Early intervention offers a diversity of family and child programs. Trained professionals provide screening and assessments of the child as well as a long list of developmental services. Common services include the following interventions: *physical therapy, occupational therapy, and speech therapy*. These services are typically provided at no cost.

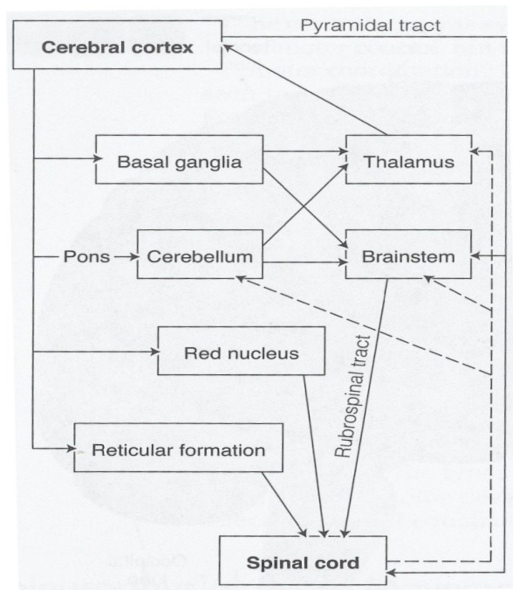
III. Theoretical Foundation of Physical and Occupational Therapy

Both physical and occupational therapists frequently focus their treatment on facilitation of normal movement patterns as well as posture and refined motor control. Treatment is grounded in the theory of neural plasticity. Theoretically, the brain has the ability to reorganize neural pathways based on new experiences. This concept is referred to as "neural plasticity" and is the foundation for delivery of early developmental therapy.

Animal models have established that "new" experiences allow for re-organization of cortical maps. In the first few years of life, although the majority of neurons have been formed, individual experiences drive modulation of neuronal death, stabilization of synapses, axonal reorientation, and budding of axonal dendrites.[38] Thus, early interventional services have the potential to reorganize neuronal pathways to improve functional outcomes.

In human movement the cerebral cortex, cerebellum, basal ganglia, brain stem, and spinal cord are the main neuronal structures guiding motor control. Specifically, the frontal lobe of the cerebral cortex contains the premotor cortex, the motor cortex, and the supplemental motor cortex regions. Through complex interactions with other regions of the central nervous system, these structures guide and execute voluntary movement. The cerebellum assists with coordination and timing of movement while the basal ganglia modulates higher and lower brain functions.[39]

Figure 1. Motor Pathways



Reprinted from: PIEK JP. Infant Motor Development. In: Wright JP, ed. Champaign, IL: Human Kinetics, 2006.

Research suggests developmental and environmental stimulation can affect neuronal cell formation as well as organization of synaptic connections in the brain over the lifespan.[40-43] Merzenich and colleagues investigated the plasticity of the brain by mapping the topographical orientation of the fingers in the cortex. When two fingers were amputated, the location of those digits was eventually taken over by the palm and adjacent fingers. Furthermore, when monkeys were taught to pick up food with the tips of their fingers the corresponding cortical brain region enlarged.[44] These studies suggest that structures are not "hard wired" for a given function. *Therefore, we hypothesize that environmental/developmental stimulation of low birth weight children has the potential to alter motor performance.*

IV. *Present State of the Literature*

Early intervention and Motor Outcomes among LBW children

Services delivered to improve developmental outcomes, specifically motor ability, typically include two theoretical types of programs. Programs, in general, either deliver treatment directly to the child to facilitate motor milestone attainment or provide education to the families to facilitate infant child interaction. Services delivered directly to the child to improve motor outcomes are typically provided by licensed physical or occupational therapists. Facilitation of gross motor outcomes falls within the scope of practice for both disciplines. These services are referred to as "interventional therapy services".

Over the past twenty years, a substantial number of randomized controlled trials evaluated the effect of early intervention services on development among low birth weight children. These studies included direct interventional therapy, mother child interventions, and developmental education curricula. To date, two large trials reported positive effects of early interventions services on neurodevelopment.[45,46]

Presently in the United States, the Infant Health and Development Program is the largest randomized controlled trial. This program was implemented in the mid-1980s to evaluate the effect of early intervention on cognitive competence, behavioral competence, and health status of low birth weight preterm infants. Researchers randomized approximately 1,000 infants to receive learning activities in a child developmental center and in the home. Parents also participated in regular support group meetings. At thirty six months, children in the intervention group had significantly higher mean IQ scores.[45] Moreover, the effects of the educational curricula varied by maternal and infant characteristics. Among the heavier babies, Stanford Binet scores were significantly higher in the intervention group compared to the control group. Researchers also observed a small statistically significant behavioral advantage among babies with less educated mothers. The results did not show any difference in serious medical conditions between the two groups.[45]

In the United Kingdom, the Avon Premature Infant Project evaluated the effect of developmental education and parent advice on neurodevelopment. Three hundred premature infants less than 33 weeks were randomly assigned to developmental education, social support, or usual care. These interventions were implemented from hospital discharge until age two. At twenty four months the results showed, for all three groups, no statistically significant difference in the mean Griffith Mental Developmental score. However, the results did show a statistically significant interaction by birth weight and presence of brain lesions. Among, children with abnormal cranial ultrasounds (hemorrhagic or ischemic lesions) or who were very low birth weight (<1251 grams), the intervention had a statistically significant beneficial effect. This effect was not observed among heavier infants.[46]

These two large trials describe the effect of parent education programs on overall development in low birth weight toddlers. Although physical and occupational therapy services were not directly implemented in these trials, the results suggest that early intervention services delivered before the age of two may have positive effects on development. *The period between birth and age two may be a critical time window for neuroplasticity among subgroups of low birth weight children.*

Several smaller randomized trials (N≈50) evaluated the effect of similar parent child programs on gross motor development among low birth weight toddlers. [47,48] Both interventions involved mother child interactions where the parent promoted perceptual, cognitive, and motor skills. At approximately three years of age, children of parents who were trained in these developmental techniques had higher mean Bayley scores in eye hand coordination, personal, social, and practical reasoning skills.[48] Furthermore, children who received parent led motor control techniques has a statistically significant improvement in object control compared to children with usual care (Table 1).[47]

Physical or Occupational Therapy and Motor Outcomes among LBW children

In the published literature, there are multiple small randomized trials that specifically evaluated the effect of physical, occupational, or physiotherapy services on motor outcomes in low birth weight children.[11,12,49-57] These trials began in the mid nineteen eighties with the majority of these studies evaluating the effect of treatment before one year of age with follow up at two years of age.

The randomized trials published to date had small sample sizes of approximately 150 children. The largest randomized trial evaluating the effect of pediatric physical therapy included one hundred and seventy six babies less than 32 weeks or less than 1,500 grams. The smallest trial included only nineteen infants. At study entry, researchers often assigned infants a neurodevelopmental score and stratified infants into three categories: normal, at risk, and neurologically impaired. Infants were then randomized to either interventional physical/occupational therapy or normal care within each stratum. Therapy was usually initiated within three months of age (chronological age) and treatment continued through twelve months. Motor skills were typically assessed using the following norm referenced standardized instruments: Bayley Scales of Infant Development, and the Griffiths Mental Developmental Scale. The frequency of treatment across these studies was highly variable and ranged from one sixty minute session per week to one treatment session per month and anywhere in between.[11,49-53,57]

Overall, several studies found no statistically significant difference in the Griffith developmental quotient or locomotor subscale scores at one year for "at risk" or "normal" low birth weight infants.[12,49,50] However, between one and two years of age, the effect of interventional therapy appeared to vary by specific infant characteristics. Overall, at sixteen months of age, children receiving the intervention who were born with low birth weight did not demonstrate a statically significant improvement in their mean Bayley psychomotor score. Yet, infants who were less than 1500 grams at birth showed greater gains in their Bayley mental score than infants between 1500 and 2000 grams at birth.[51] At age two, children in this population who received six to eight sessions of pediatric physical therapy had an increase of 6.4 points on the Bayley psychomotor scale after adjusting for perinatal and background variables. Moreover, in subgroup analyses, these researchers found improvements in both motor and mental outcomes among children with a history of bronchopulmonary dysplasia and biological and social risk factors for preterm birth.[57]

One study in a Turkish population evaluated the effect of physical therapy intervention on actual age of motor milestone achievement. These authors recruited a small sample of 160 infants less than 34 weeks gestation and 2,000 grams from the Hacettepe University Hospital in Turkey. Infants were stratified into two groups: those with perinatal hypoxia or abnormal neurosonography, and infants without any risk other than prematurity. Researchers randomized the "low risk" infants into an interventional pediatric therapy group (N=78) or a control group (N=76). Children received approximately 17 therapy sessions between birth and two years of age. Over the course of follow up, researchers reported no statistically significant differences in age of motor milestone achievement between these two groups.[55]

The literature evaluating the efficacy of occupational and physical therapy services on motor performance among preschool age low birth weight children is sparse. In a meta-analysis of preschool age children of normal birth weight, interventions that promoted motor skills appeared to improve early childhood object control and locomotor skills.[13] However, two small randomized studies evaluating the effect of physical therapy and occupational therapy, delivered before twelve months of age did not find a statistically significant difference in school age motor scores compared to controls.[11,12] These studies evaluated therapy services in a small population (< 100 children) of VLBW children without neurological problems (Table 2).

Presently, in the published literature, the majority of studies evaluated the effect of interventional therapy services delivered during the first year of life. The focus of treatment during this developmental window was most likely on fluidity of movement and independent transitions with pre ambulatory skills including head control, independent sitting, crawling, and walking with support. However, research evaluating the effect of therapy between one and two years of age and motor ability in later childhood is extremely limited. Therapy delivered during the toddler years may include facilitation of foundational skills of ambulation, strengthening, coordination, and balance. This type of treatment may more directly carry over into improvement in more complex movement patterns.

V. Estimating Effects in Observational Data

Observational data are publically available to evaluate the effect of interventional therapy on school age motor performance. However, use of these data is complicated by differences in measured and unmeasured characteristics that are independently associated with motor performance.

Propensity scores are commonly used to control for confounding in observational studies. The propensity score is the predicted probability of receiving treatment given a set of measured confounders.[14] Subjects with the same probability of receiving treatment have similar values of measured characteristics which are independently associated with the outcome. Once one conditions on the propensity score, any difference in the distribution of measured covariates between treatment groups should be from chance alone.

Currently, there are few studies to guide the researcher on how to estimate the propensity score. Logistic regression is commonly used to estimate the propensity score; however the model is subject to several assumptions. The model assumes a linear relationship between continuous and ordinal variables and the logit of the dependent variable. Moreover one must consider the joint effect of variables. Inclusion of only main effects may misspecify the model and the resulting effect estimate may be biased. Yet, it appears in the published literature, that few researchers consider interactions or the functional form of the variable.[58]

Regression tree based methods, including Bagging and Random Forest classification (RFC), are non-parametric methods derived from learning based algorithms which offer alternative strategies for generating predicted probabilities of treatment. The methods use a series of classification trees to estimate the average probability of membership in a given class. These techniques have been suggested to have improved predictive accuracy when compared to classical statistical techniques.[59] For example, in simulation studies, regardless of non-linearity or non additivity, random forest performed well in terms of covariate balance between treatment groups and may result in further reduction in bias of the effect estimate when compared to traditional logistic regression.[17,18]

The large nationally representative cohort proposed for this study, ECLS - B, provides nationally representative data regarding receipt of interventional therapy between nine months and two years of age and motor development through kindergarten. These data will allow us to explore the association between receipt of services and motor development during the preschool and school age developmental periods.

Chapter 2

METHODS

Data Source

We will address the specific aims of this study using existing data from the United States Early Childhood Longitudinal Study Birth Cohort.

I. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)

Overview

The ECLS- B is a nationally representative longitudinal cohort of children born in the year 2001 who were followed through kindergarten. The study was sponsored by the United States Department of Education and the National Center for Education Statistics. Researchers obtained information on children's physical, social, emotional, and cognitive development as well as health and education over the child's early developmental years.

Research Design and Sampling

The ECLS-B is a longitudinal cohort study which followed children from nine months of age through kindergarten. Researchers collected data by both questionnaire and direct assessment at four time points: nine months (2001-2002), age two (2003-2004), age four (preschool: 2005), and kindergarten (2006-2007).

In 2001, the study team randomly sampled 13,500 newborns from United States birth certificates. Asian, Pacific Islander, and Chinese children were oversampled. Researchers also oversampled twins as well children who were born low birth weight. At the nine month time point, approximately 10,700 children and their parents participated in data collection. At two years of age, approximately 9,850 children remained in the study.

Data Collection

Researchers collected information on APGAR test scores, parental background, and other child health information from the birth certificate records. Information on a child's cognitive, social, emotional, and physical development were collected from children's families, child care providers, and teachers.

Questionnaires:

At each wave of data collection families completed two questionnaires: the parent interview, and the resident father questionnaire. The parent questionnaire ascertained information regarding demographics, family structure, child development, the home environment, parent attributes and expectations, child care arrangements, child health, family health, marital history, social support, community support, respondent information, spouse information, and information on the nonresident father.

The resident father questionnaire ascertained information on education, employment history, childbearing, marital partner history, separations from the child, parenting practices, knowledge of child development, prenatal experiences, and home involvement.

Childcare providers also completed information on center services and staffing.

Direct Assessments:

Researchers completed direct child assessments in the area of cognitive performance, socio-emotional development, and physical performance over the follow up period. The study team assessed physical growth and motor development with two standardized assessments: The Bayley Short Form Research Edition, and items from the Bruininks-Oseretsky Test of Motor Proficiency (Bruininks) and the Movement Assessment Battery for Children (MABC). They also obtained direct measures of weight, height, arm circumference, and head circumference. The Bayley was administered at nine months and two years of age. Items from the Bruininks and MABC were administered at preschool and kindergarten.

Low Birth Weight Cohort

The ECLS-B oversampled children who were less than 2,500 grams. At baseline, the cohort included approximately 1,650[a] children who were low birth

[a] Numbers have been rounded to the nearest 50 according to the data use agreement with ECLS-B

weight and 1,150 children who were very low birth weight. Approximately 2,900 children were preterm. Approximately ninety one percent of families completed parent interviews at two years (N≈2,500), preschool, and kindergarten (N≈1,900) waves of data collection.

Exposure Assessment

We will define interventional therapy services, as those therapies implemented to facilitate motor development in childhood. Children frequently receive physical and occupational therapy services concurrently. Facilitation of gross motor skills falls within the scope of practice for both disciplines. Yet, there may be some divergence in the tactics that each discipline implements to improve motor control. However, it is difficult to isolate the individual effects of each service. Therefore, we will consider interventional therapy as children who received *physical* and or *occupational therapy*.

The ECLS-B cohort assessed exposure to therapy services when the child was 9 months of age, at two years of age, during preschool, and at entry into kindergarten.

Researchers asked the parents at 9 months: “For each service, please tell me if child or your family received this service to help with special needs”. At 24 months, researchers asked “Since your last interview does your child receive therapy services? On the preschool questionnaire, families were asked “Since the age of two has your child received speech, physical therapy, occupational therapy, or vision services”.

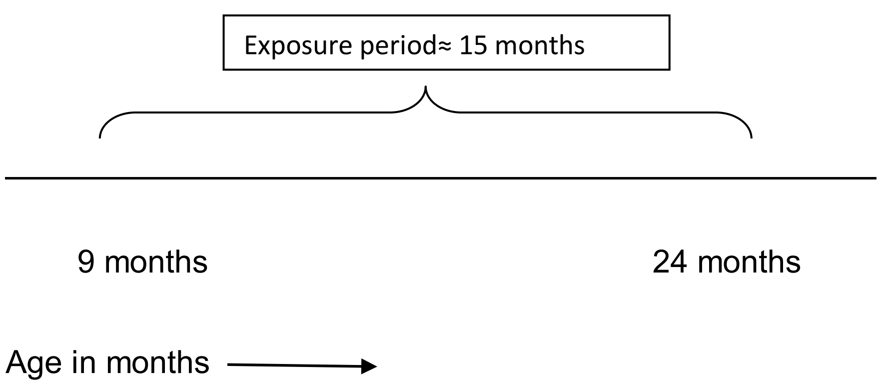
We will use the following criteria to define the exposure:

1) If the child received physical and or occupational therapy services any time between 9 and 24 months, they will be considered exposed. We will model this variable as a dichotomous outcome.

Table 3 provides the distribution of services that low birth weight children received at nine months.

Figure 2. Exposure definition ECLS-B

"For each service, please tell me if child or your family received this service to help with special needs."



Outcome Assessment

The ECLS-B assessed motor development using both standardized assessments and parent report of developmental milestones.

We will assess the motoric ability of this low birth weight population using standardized assessments and caregiver report of developmental milestones.

Standardized assessments:

a) The *Bayley Scales II of Infant Development* is one of the most widely used standardized measures to assess cognitive and motor performance among children 0-42 months of age. The instrument provides standard scores for two indices: the Mental Developmental Index (MDI) and the Psychomotor Developmental Index (PDI). Composite scores are calculated for each index and compared to "typically" developing children of the same age. In the late 1980's , the instrument was standardized using 1,700 US children born at 36-42 weeks who were normal weight for gestational age and without medical conditions or disabilities . The concurrent validity of the Bayley Gross Motor Scale and the Peabody Developmental Motor Scale show a high correlation (r=0.83).[60] Researchers evaluated child developmental skills at one month intervals between 1-42 months of age. The assessment provides a standard score for the Psychomotor Index with a mean of 100.4 and a standard deviation of 16.2. The lowest Psychomotor Developmental Index standard score on this instrument is 50.

Investigators used the Bayley Short Form Research (BSF-R) Edition Motor Scale to assess gross motor performance at nine month and two years of age. The BSF-R includes a subset of the BSID II. Researchers used Item Response Theory (IRT) modeling to select items that represented all constructs of the BSID II. The study team established that the BSF-R could be used in place of the BSID II to measure developmental performance. The scaled scores for this measure represent the same metric as the BSID II.[61]

Data from the Bayley Short Form are available at both 9 months and age 2. We will use this instrument as an indicator of baseline functional ability during the nine month assessment period.

b) *Bruininks -Oseretsky Test of Motor Performance:* The Bruininks is a norm referenced test designed to assess both gross and fine motor functioning among children between the ages of 4 1/2 to 14 1/2. The test was standardized using 765 children in the United States with standard scores and percentile ranks by age grouping. The full test includes 46 items that are divided into the following subtests: running speed, agility, balance, bilateral coordination, strength, upper limb coordination, response speed, visual motor control, and upper limb speed and dexterity.[62] Composite scores can be generated separately for gross motor and fine motor sections. Studies reported an intraclass correlation coefficient of 0.85 for gross motor subscales of the gross motor composite score. Factor analyses suggest this instrument has poor construct validity in discriminating between fine and gross motor ability. The instrument offers a valid measure of general motor proficiency.[63]

c) *Movement Assessment Battery for Children (MABC)*

The MABC is a widely used standardized assessment to evaluate motor impairments in children between the ages of 3 and 17. The assessment is frequently used by educators, physical therapists, and occupational therapists to identify deficits in motor impairments. The assessment evaluates gross motor skills in the following areas: manual dexterity, aiming, catching, and balance. This instrument is frequently used to identify children with DCD.[64]

The ECLS-B did not administer the Bruininks or MABC in their entirety. Rather, investigators chose select items from these assessments. Children to complete the following skills at each time point: skipping eight consecutive steps, hopping on one foot five times, walking backwards six steps on a taped line, standing on one foot for ten seconds, and jumping forward from a standing position.

Children received one trial to complete the skipping and walking backwards items and three trials to complete the balance and hopping items. Investigators scored the items on a pass fail basis. For the balance and hopping items, investigators also recorded the greatest number of hops as well as the greatest number of seconds the child balanced on one leg across all trials. For the jumping task, each child received two trials, and investigators reported the greatest distance the child jumped in inches.

Investigators also asked children to catch a thrown bean bag. The score for this skill was the number of times that the child caught the bean bag. The ceiling for this item was low, at five tosses.

We will model the association between receipt of interventional physical and occupational therapy and each individual preschool motor item. We will model the items scored on a pass fail basis as a dichotomous outcome. We will model jumping distance as a continuous outcome. Due to the low ceiling on the catching skill, we will not include this item in our analysis.

We will calculate the change score in jumping distance between preschool and kindergarten.

Developmental Milestones

Parents reported the age at which their child completed gross motor milestones at two time points: 9 months and two years.

- 9 months: parents reported the age at which their child sat independently, crawled on hands and knees, pulled to stand, and cruised

- 24 months: the parents reported the age when their child started walking up the stairs alone

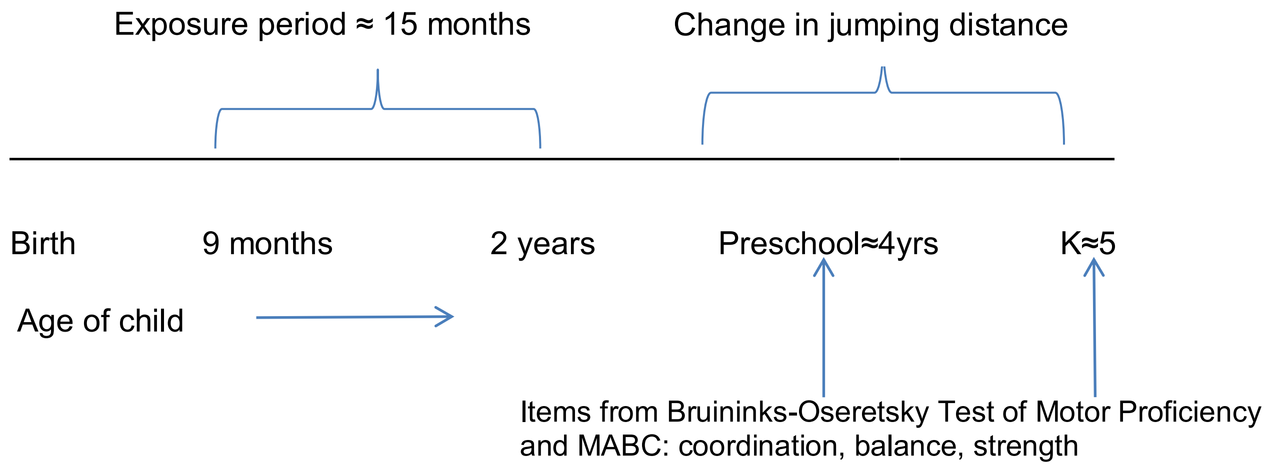
At nine and 24 months, ECLS-B included all children in the one on one assessment. When children reach preschool and kindergarten, children in wheelchairs did not participate in gross motor assessments. Children, who used an assistive device, were allowed to use that device during the assessment.

We will use caregiver report of developmental milestones to determine if the child is attaining typical developmental milestones during age appropriate time periods.

Method

Study Design: We will conduct a cohort study to estimate the effect of physical or occupational therapy on preschool motor coordination among children born LBW.

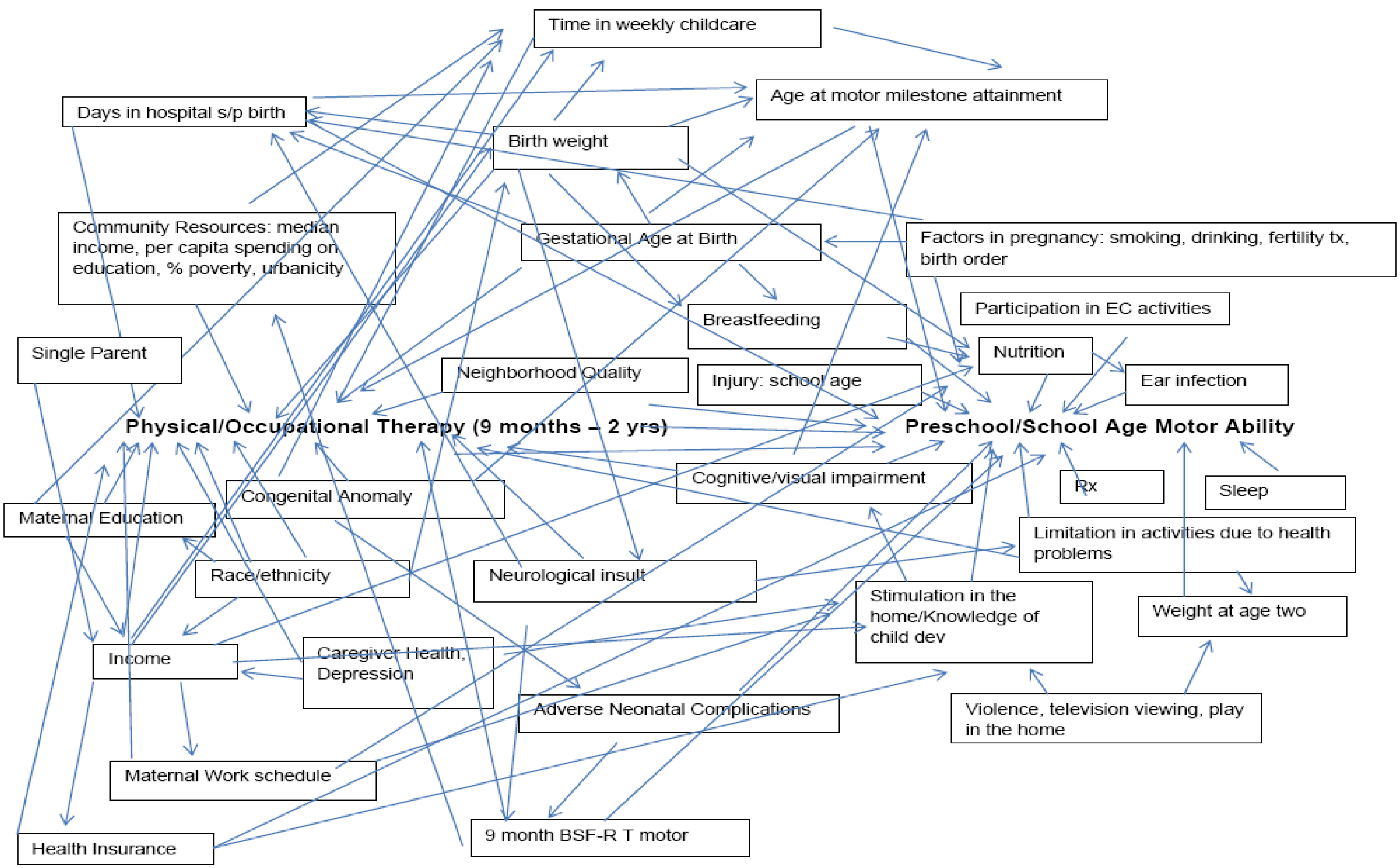
Figure 3: Study Design



Directed Acyclic Graph:

This graph represents unidirectional causation between the therapy services, preschool motor ability, and confounding variables (Figure 4).[65]

Figure 4. Directed Acyclic Graph



Analysis Plan

I. General Overview

We will estimate the effect of interventional physical occupational therapy services on preschool motor development using data from the ECLS-B.

Children who receive therapy services in this low birth weight population are a heterogeneous group. Naturally, children will have different “propensities” to get services based upon a host of demographic, medical factors, and functional ability. Therefore, in these observational data, we may have confounding between therapy and motor development by severity of the child’s medical condition and or functional level. Children that are sicker are more likely to get services than those children who are healthier.

II. Estimating the average causal effect in observational data

We will use two general approaches to estimate the average treatment effect of therapy on preschool motor coordination. First we will use a standard logistic regression model while controlling for confounding. Second we will use a propensity score approach. We will estimate the predicted probability of treatment using both tree based methods and standard logistic regression. We will then create inverse probability of treatment weights with each method to then estimate the average treatment effect of therapy services on preschool motor coordination.

We will use measured confounders at nine months to control for confounding of the association between receipt of therapy services and preschool motor performance. These methods are based on the assumption of no unmeasured confounders in the data. Using a priori knowledge and a directed acyclic graph, we will include confounders associated with both receipt of therapy services and motor outcomes in the propensity score model. Moreover, to decrease bias and improve precision of the effect estimate, we will also include those covariates that are associated with motor development.[66]

A. Standard Model:

We will generate separate models to estimate the effect of interventional physical or occupational therapy services on preschool motor ability. We will use a logistic regression model, while controlling for relevant confounders, to estimate the average effect of therapy on the ability to hop five times independently, to skip eight consecutive steps, to maintain single leg stance for ten seconds, and to walk backwards six steps on a taped line.

$$Logit\ (Pr(Y=1) = B_0+B_1X_1+B_2X_2+B_iX_i+e$$

Using a linear regression model, we will model the estimated effect of therapy services on preschool jumping distance and on the change in jumping distance (preschool to kindergarten), while controlling for confounders.

$$Y= B_0+B_1X_1+B_2X_2+e$$

We will consider the functional form of the covariate and evaluate the assumption of linearity between continuous and ordinal variables and our preschool outcomes for both the linear and logistic models.

We will consider modification of the effect estimate by maternal social characteristics (maternal education). These variables are based on interaction effects reported in the literature. We will use the Breslow Day test of homogeneity to determine if there is heterogeneity of the odds ratio across strata of the covariate. Mostly likely we will encounter small cell sizes when the main effect is stratified by level of each interaction term. We will use a less stringent p value, $p<0.10$, since the power to detect interactions is often low.

B. Inverse probability of treatment weights (IPW)

We will create inverse probability of treatment weights to estimate the average treatment effect. These weights will create a pseudo population with the distribution of the covariates reflective of the combined sample. The weights are calculated from the propensity score, the predicted probability of treatment given a set of covariates.

The weights are as follows:

a) 1/(propensity score): if the child received therapy

b) 1/(1-propensity score): if the child did not receive therapy

Children with a low probability of receiving treatment and actually received therapy receive a large weight, and children with a low probability of treatment and didn't get therapy receive a small weight.

We will stabilize the inverse probability of treatment weights by multiplying the IPW weight by the marginal prevalence of the treatment that they actually received. Stabilizing the weights "normalizes" the range of the probabilities and increases the efficiency of the analysis. This prevents just a few people from contributing most of the observations in the pseudo population. Stabilizing the weights centers the weights around 1.0.[67]

We will estimate the average treatment effect of interventional therapy delivered between 9 months and age two on preschool motor ability using propensity scores weights from each estimation method. We will generate separate models for each preschool motor item as described above. For each weighted model, we will control for residual confounding.

Estimating the Propensity Score:

We will estimate the probability of a child being exposed to interventional therapy services between nine months and age two using four methods: *logistic regression, a single classification tree, random forest classification, and bagging*. The propensity score provides a weighted summary of the covariates. Theoretically, when conditioning on this score, the distribution of measured covariates should be similar between treatment groups. Thus, the variation in the covariates between the groups should be from chance alone.

1. Logistic Regression: We will estimate the propensity for early childhood therapy with a logistic regression model. This model assumes linearity between covariates and the receipt of therapy.

$$\text{Logit (Therapy=1)} = B_0 + B_1X_1 + B_2X_2 + B_2X_3 \ldots\ldots + B_pX_p + e$$

2. A Non-Parametric Approach to Estimating the Propensity Score

Recursive partitioning

Recursive partitioning is a non-parametric classification and regression tree method which is commonly used in clinical medicine and genetics. This approach analyzes large numbers of predictor variables and complex interactions to create regression trees. The method partitions the data into subgroups which show the greatest heterogeneity with respect to the outcome. The method is objective and data driven, therefore the groupings will be automatically generated by the software package. The subgroups are objective and mutually exclusive.

During each stage of partitioning, observations with similar outcome responses are grouped. Unlike linear regression where information is combined linearly, here recursive partitioning considers both nonlinear associations and multiple splits of the same variable. This method may be similar to stepwise regression where candidate variables are entered into the model one at a time, however with recursive partitioning; only those interactions which are used to grow the tree are used to fit the data.[68]

The tree is grown according to the concept of “impurity reduction”. With each split in the building process, the association between the “daughter nodes” and the outcome are more homogenous compared to the previous parent nodes. As the tree grows, variables that are more strongly associated with the outcomes are chosen to split. Many classification trees rely on p values for tests of association to determined cut points.[68]

This splitting continues until a “stop” point is set. Criteria for a “stop” point may include a threshold for the minimum number of observations in the node or a threshold for the minimum change in the impurity measure.[68]

Despite the popularity of this data mining method, results from a single classification tree are highly variable and are sensitive to the arrangement of the data. For example, the rank of each variable in the classification tree as well as the cut point of the variable is strongly dependent upon the distribution of observations in the data. With small changes in the data structure, the order of variable selection or the cut point of the variable may change resulting in an alternative tree structure.[68]

Bagging and Random Forest Classification

Both bagging and random forest classification are tree based methods derived from machine learning theory which aggregate estimates over multiple individual trees to improve the predictive performance of the algorithm. Bagging randomly draws a series of bootstrap samples from the data, and creates individual classification trees for each sample. With each of these bootstrap samples, the data will vary slightly from the previous sample. Furthermore, each individual tree may then vary, perhaps substantially, from the previous tree. The algorithm then aggregates the predicted probability of class membership over the series of classification trees.[69]

Random forest classification utilizes this same bootstrap method. However, random forest adds an additional level of variability to the algorithm. During construction of the individual classification trees, a random sample of predictor variables is chosen to split the data at each node. Therefore, each individual tree is even more diverse compared to the trees from bagging alone.[68]

Although individual classification trees are inherently unstable, bagging and random forest classification have been shown to produce robust estimates. In both empirical and simulation studies, estimates aggregated over a series of classification trees, show improvements in prediction accuracy when compared to a single classification tree.[70-73] Bagging is suggested to equalize the influence of given observations in the data. Thus, data points which strongly influence the classification algorithm are downweighted.[68] Furthermore, the additional level of randomness introduced by random forest classification creates additional diversity between trees with a lower upper bound of error.[69] Overall, these methods produce a more robust final estimate with decreased variability.[69]

We will generate predicted probabilities of class membership into interventional therapy between nine months and age two using logistic regression, random forest classification, bagging, and a single classification tree.

These analyses will be performed using the R statistical platform. We will use the RandomForest (random forest), Ipred (bagging), and Tree (single classification tree) packages.   All categorical variables will be encoded as “factors” in the R environment. This transformation ensures that the R software recognizes these variables as categorical responses.

We will check the sensitivity of the error rate to our chosen parameters by allowing the number of trees to vary between 250 and 1,000 and the number of randomly chosen variables to vary between 2 to 7. The error rate for the algorithm is generated from the 33% of the data remaining that was not used to form the classification trees. For example, with each bootstrap sample, the remaining data (≈33%) not in the sample is entered into the classification tree. The error in these out of bag predictions is collected over the series of trees to determine the final error rate over the forest. The error rate appears to be accurate if the predicted probabilities of class membership are aggregated across a sufficient number of trees. However, if the number of trees are too few, then the error rate may be upwardly biased.[74] The algorithm may therefore be a better predictor of the outcome than suggested by the error rate.

We will assess balance of each method used to generate the propensity score by calculating the standardized difference of the weighted confounding variables between the treatment groups.

Standardized differences represent the differences between the means by therapy status in units of standard deviations. The estimate is calculated as

$d = (\bar{x}_{therapy} - \bar{x}_{no\ therapy}) / \text{sqrt}((S^2_{therapy} + S^2_{no\ therapy})/2)$. [75] Although there is no standard criterion to determine balance between treatment groups, researchers suggest a standardized difference of <0.10.[76-78]

<u>Limitations of these approaches</u>:

These approaches attempt to control for confounding due to non-randomization assignment of treatment. Propensity scores can balance the distribution of the covariates between the groups, so when conditioning on the propensity score, there is no longer unequal distribution of covariates between groups. Therefore the association is no longer confounded. However, this is dependent on the variables that are measured in the data as well as the variables that are included in the propensity score model. We still may have unmeasured confounding.

In addition, the propensity score model may be misspecified if the analyst does not consider the functional form of the confounders or higher order effects. We will consider two methods from machine learning theory that are free from these parametric assumptions.

Sources of Bias

1) Attrition and Selection Bias:

Estimation of the predicted probability of treatment requires complete data on covariates used to generate the predicted probability of treatment. Children who are missing data to generate the probability of treatment may differ from the children who have complete data.

Attrition between twenty four months and preschool/kindergarten is also a concern, as children who have preschool outcome data may differ, in meaningful ways, from children who do not have data on preschool motor outcomes. These data were collected as a larger study evaluating early childhood health, development, and education so we do not anticipate that attrition would be related to children no longer receiving therapy services. Therefore, there would not be differential loss to follow up based on children doing "better" as a result of therapy.

The proportion of children with preschool follow up data was similar among children who did (87%) and did not (84%) receive therapy services between 9 months and 24 months.

Handling of missing data:

Although the assumption can't be tested in the data, under the missing at random assumption, we will consider proc Iveware to impute the missing data.

Proc Iveware imputes the missing values using multivariate sequential regression. We will impute the missing values for the raw data prior to making transformations or collapsing variables. We will impute five data sets, and run all analyses in each of the imputed data sets. The effect estimates will be averaged across imputations. The standard errors will be calculated to account for variation both within and between imputations. We will calculate the standard errors across imputations using Rubin's Rule.[79]

Missing data and recursive partitioning:

Compared to standard regression methods, tree based approaches do not completely discard an observation with missing data. The observations that have missing data in a variable which is being partitioned are not included in that split. However, that observation will be included in other computations of variables where that observation has a data point. Random forest classification and bagging algorithms presently do not handle missing data. We will use complete covariate data to generate the predicted probability of treatment using these algorithms.

Sensitivity Analysis:

There is the potential for unobserved confounding of the effect estimate due to unmeasured variables that are not included in the propensity score.

We will examine the sensitivity of our effect estimates by restricting our sample to those children with overlapping propensity scores. From our sample with overlapping propensity scores, we will also exclude children who were treated most contrary to prediction. For example, we will exclude children who received therapy but had a propensity for treatment lower than the 1st percentile and children who did not receive therapy but had a propensity for treatment greater than the 99th percentile. We will also trim the sample using the 2.5th and 97.5th percentile cut points for the treated and untreated children respectively.[80]

Simulation studies show, assuming a uniform effect of treatment, that asymmetric trimming of the propensity score leads to a reduction in bias in the presence of unmeasured confounding.[80]

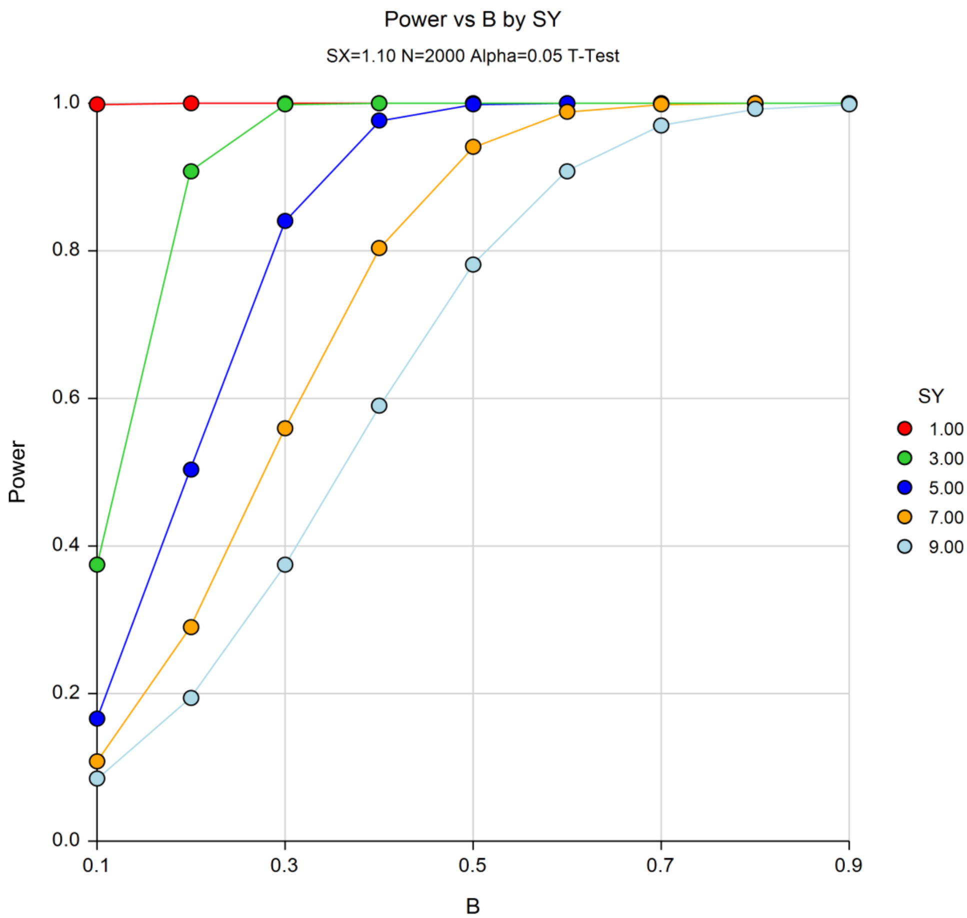
Power Calculations

The ECLS-B cohort included approximately 3000 low birth weight children. We estimated the standard deviation of gross motor performance among school age children to range between 1 to 10.[81] At 80% power, we can expect to detect a change of 0.30 to 0.50 in the mean gross motor score.

ECLS-B: Logistic regression power analysis with imputed sample: early childhood

Power	Approximate Sample Size	Proportion receiving PT/OT	Probability of Motor Delay	Odds Ratio	Alpha
0.09709	1255	60	0.2	1.1	0.05
0.46585	1255	60	0.2	1.3	0.05
0.84255	1255	60	0.2	1.5	0.05
0.97707	1255	60	0.2	1.7	0.05
0.99817	1255	60	0.2	1.9	0.05

Figure 5. ECLS-B Linear regression power analysis



Chapter 3

PRELIMINARY ANALYSES

VARIABLE DESCRIPTION AND CODING

EXPOSURE:

At two years of age, researchers asked families of participating children if their child was receiving related services between nine months of age and age two. Families reported the type of therapy that the child received during this time frame as well as the amount of therapy the child received per month. Families reported total amount of therapy across all disciplines. We defined receipt of either physical or occupational therapy between nine months and age two as a dichotomous variable.

OUTCOME:

Researchers administered select items from the Bruininks Oseretsky Test of Motor Proficiency and the Movement Assessment Battery for Children to assess preschool and school age motor ability. The child completed the following tasks at preschool and kindergarten: skipping eight consecutive steps, hopping independently five times, maintaining single leg stance for ten seconds, walking backwards eight consecutive steps on a taped line, and performing a standing broad jump.

COVARIATES

Continuous Measures:

1. Center for Epidemiologic Studies-Depression Scale. The items were scored according the ECLS-B manual and then summed for each individual who had fewer than four missing responses. We coded subjects with four or more missing items as missing on the composite CES D variable.

2. Home Observation for Measurement of the Environment (HOME): Measures the extent of quality and amount of child stimulation in the home environment. Researchers administered a subset of 8 items from HOME-SF which included items from the following subscales: responsivity, acceptance, involvement, learning material, organization subscales.

   These items are dichotomous answers with either positive or negative scores. These variables were recoded as “yes” =1 and “no”=0. Not observed were considered “missing”. The variables were summed to create the final HOME score at 9 months. Select questions from the HOME scale included: parent spontaneously vocalizes to the child, parent verbally response to the child’s vocalization, parent caresses or kisses the child at least once, parent neither slaps nor spanks child during visit, parent does not interfere with or restrict child more than three times during the visit, parent provides toys to the child to play with during visit, parent keeps child in visual range, child’s play environment is safe. We calculated the HOME score by taking the average across all items and multiplying this value by the number of question completed.

3. Knowledge of Infant Development Inventory (KIDI): 11 items from the KIDI designed to assess knowledge of parental practices, developmental processes, and infant norms of behavior. These 11 items were selected from 75 items on the KIDI questionnaire that the authors of the instrument recommend as the most successful items in differentiating high versus low parenting knowledge. The questions describe typical infant behavior or parenting that would affect infant growth and behavior. Parental responses include: “agree”, “disagree”, “not sure”. These questions are a measure of an individual child’s development. We calculated the KIDI score by taking the average across all items and multiplying this value by the number of question completed.

4. Birth weight: all children in the cohort weighed less than 2500 grams. Birth weight was ascertained from the birth certificate record.

5. Number of siblings: This variable is continuous and indicates the total number of siblings either full, step, adoptive, or foster that lived in the household with the child at the nine month assessment

6. Gestational age at birth: gestational age was ascertained from the child’s birth certificate record

7. 5 Minute APGAR scores: researchers ascertained the five minute APGAR scores from the birth certificate record

8. Length of hospital stay: at the nine month assessment families reported the number of days that the child stayed in the hospital at birth due to medical problems

9. Length of NICU stay: at the nine month assessment families reported the number of days since birth that the child stayed in the NICU

10. 9 month BSF-R Motor T Scores: standardized t scores of motor performance indicating the child's ability relative to other children the same age. The scores are norm references with a mean of 50 and a standard deviation of 10. Standardized T scores are adjusted for gestational age.

11. SES scale: this is a continuous measure for the composite of socioeconomic status which ranges from -2.10 to 2.25. The composite is the average of up to five measures: mother's education, father's education, mother's occupation, father's occupation, household income.

12. Hours per week in childcare: at the nine month assessment families reported the number of hours per week that their child spent in childcare.

13. Age of motor milestone attainment: Researchers asked parents to report whether their child could perform the behavior and when the child was first able to perform the skill. On the nine month assessment parents reported when the child first performed the following skills: sit alone and steady without support, crawl on hands and knees, pull to a standing position, and first walked holding onto something (cruising). Researchers derived these items

from the Minnesota Child Development Inventory. The appropriate age range for this measure is from birth through age six.

Categorical Variables

1. Parental education: We collapsed parental education into the following five categories: less than a high school education, high school, technical training/some college, college degree, and graduate/professional training. We coded education as a series of indicator variables.

2. Caregiver Health: caregiver's reported their current health status during the nine month follow up assessment. Responses included excellent, very good, good, fair, and poor. We coded these responses as a series of indicator variables.

3. Injury of the child: how often the child was seen for an injury by a professional since the child has lived in the home

4. Child Health Condition: On the nine month questionnaire, researchers asked the caregiver if a doctor ever told them that their child had a health condition. This question is coded as a series of dichotomous variables with 1=yes and 0=no. The health categories are as follows: visual deficit, cleft palate, heart defect, congenital anomaly affecting motor skills, failure to thrive, difficulty with mobility, and other special needs. Turner's syndrome, Spina Bifida, and Downs Syndrome were included under the category of congenital anomaly.

5. Single parent status: researchers asked respondents if a spouse or partner lives in the household. This variable is coded as a dichotomous variable where 1=”yes” and 0=”No”

6. Help or advice with childcare: Researchers asked the primary caregiver, who you would ask for care and advice about your child. This variable was collapsed into a dichotomous variable where: 1=yes, the caregiver received advice, and 0= no, the caregiver had no one to ask for advice.

7. Health Insurance: at the nine month assessment parents reported whether or not the child was covered by health insurance as well as the type of insurance plan. Researchers asked the respondent about the following health insurance plans: private, Medicaid, SCHIP, military, Indian Health Service, and other government programs (Medicare, State sponsored health plan). 1=yes and 0 = no. This variable is coded as a series of dichotomous variables.

8. Race/Ethnicity: We collapsed race into the following categories: White Non-Hispanic, African American, Hispanic, Asian, and Other. We coded this variable as a series of indicator variables.

9. 9 month Work Schedule: Mothers were asked whether they were currently working as well as their work schedule. We collapsed the responses into the following categories: does not work, regular daytime shift, regular evening shift, night shift, rotating shift, split shift, other shift. This variable was coded as a series of indicator variables.

10. 9 month questionnaire: Injury in the last three months: Families were asked "Since you have lived here, how many times has your child seen a doctor or other medical professional or visited a clinic or ER". The range of responses included never, once, twice, more than three times. We coded this variable as a series of disjoint indicator variables.

11. Caregiver Health at nine months: This variable is an ordinal variable ranging from excellent to poor. Researchers as the respondent on the nine month questionnaire to rate their health in general. The responses ranged from excellent to poor. We coded this variable as a series of indicator variables.

12. Urbanicity: This variable is a coded as a nominal variable with three categories: urban (inside urban area), urban (inside urban cluster), rural.

13. Region: This variable indicated the region where families resided at the time of the 9 month assessment. This variable is coded as a nominal variable with the following categories: Northeast, Midwest, South, and West.

I. PRELIMINARY ANALYSES

DESCRIPTIVE ANALYSIS: LOW BIRTH WEIGHT COHORT

In these data, approximately 3,000 children were born < 2500 grams. Nine percent of the sample received interventional physical or occupational therapy services between nine months and age two. Data on receipt of therapy were missing for 8.14% of the sample. Children who received therapy were more likely to be male (60% vs. 47%) and were born on average five weeks earlier than children who did not receive therapy (29 weeks vs. 34 weeks). Overall children in this cohort who received therapy were on average very low birth weight (1138 grams vs.1785 grams) and demonstrated a lower functional ability at nine months. Children in the treatment group were hospitalized three times (75 days vs. 24 days) as long after birth and were five times as likely not to be sitting independently (45% vs. 7%) at the nine month follow up visit compared to children who did not receive treatment. Only 23% of children in the treatment group were cruising along furniture where 63% of children in the untreated group had attained this skill (Table 4).

Exploration of Confounding

We conducted exploratory analyses to determine: variables that were associated with receipt of therapy services between nine months and age two, variables that were predictive of preschool motor ability, and the functional form of the relationship between covariates and preschool motor ability. We entered a host of candidate variables into the RandomForest package in R version 2.1 (Table 5). The RandomForest package does not support missing data values. The percent of covariates missing data were extremely small (<5%) with exception of APGAR and Bayley Motor T scores in which 15% of children were missing data for these measures. In this exploratory analysis, we used proc IVEware in SAS version 9.2 to impute the missing values.

Using the RandomForest package we generated 1500 trees using 9 variables chosen randomly to partition the data at each node. Difficulty with upper or lower extremity mobility indicated by a physician, the child's length of hospital stay after birth, and birth weight had the highest mean decrease in accuracy and were very strong predictors of therapy receipt (Figure 6). The mean decrease in accuracy is the difference in classification accuracy using the out of bag data when the variable is included and the classification accuracy when the values of the variable in the out of bag variable are permuted randomly. A higher mean decrease accuracy score indicates a variable of greater importance in prediction of receipt of early childhood therapy services. [69] Other strong predictors of therapy included the child's 9 month BSF-R Motor T score, gestational age, race, socioeconomic status, and the inability of the child to attain early developmental milestones (sitting, pulling to stand, and crawling) (Figure 7).

Table 6 describes the importance of candidate variables to predict preschool motor ability. The ECLS-B did not administer the Bruininks Oseretsky Test of Motor Proficiency or Movement Assessment Battery for Children in its entirety. To explore this association, we calculated an overall motor score from the available administered items. We took the average of the following items: skipping, walking backwards, hopping left and right, balance left and right and multiplied the average by number of items without missing data. Children with a score of one standard deviation below the sample mean were considered to have impaired preschool motor coordination.

Length of hospital stay, birth weight, gestational age, and a delay in foundation motor milestones (inability to pull to stand, crawl, and cruise) were strong predictors of preschool motor impairment in this low birth weight population. Nine month BSF-R Motor T score, socioeconomic status, amount of time the child spent in childcare were also relatively strong predictors of preschool motor impairment (Table 8).

Figures 8 through 12 provide partial dependence plots between covariates and the logit of preschool motor impairment. The graphs characterize the relationship between an individual predictor variable and the probability of preschool motor impairment from the Random Forest classification algorithm. The y axis is on the logit scale. The partial dependence plot represents the marginal effect or average trend of the variable after averaging out the effects of the other predictor variables[b] in the model.[82] The logit of the probability of preschool motor impairment decreased with increasing birth weight up through 2500 grams. There was an upward trend at the upper range of low birth weight (Figure 9). There was a monotonic linear trend between both gestational age and length of hospital stay and the logit of preschool motor impairment. The logit of the probability of a preschool motor impairment decreased with increasing gestational age and increased with length of hospital stay. The relationship between socioeconomic status and preschool motor impairment was U shaped. The relationship between 9 month BSF-R motor T score and the logit of

[b] Variable in Random Forest classification model:9 month Bayley Motor T Score, Weekly childcare hours, number of siblings, length of hospital stay, length of NICU stay, hearing deficit, failure to thrive, other special healthcare need, gestational age, ever breastfed, number of ear infections, weekly hours of television watching, neighborhood safety, maternal depression, birth weight, APGAR score, HOME score, KIDI score, region, urbanicty, health insurance, maternal work schedule, PT/OT receipt, maternal support, congenital deficit, visual deficit, difficulty with upper or lower extremity mobility, childhood injury, caregiver health, age at sitting, crawling, standing, cruising, parental education

preschool motor impairment showed a relatively linear decreasing trend through a score of 50. Yet, there was a small upward trend with BSF-R score above 60 (Figure 8). The relation between many covariates (APGAR score, KIDI score, SES level, gestational age, and nine month BSF-R Motor T score) and the logit of the probability of therapy exposure was nonlinear (Figures 13-20) (Table 7).

Building the propensity score:

In our baseline cohort there was marked heterogeneity in children's baseline functional ability. We evaluated candidate confounders for the propensity to receive therapy, among children with similar baseline functional ability or "need" for therapy.

We defined an at risk cohort of children who had similar baseline levels of functional ability. This included children who were sitting independently on the nine month assessment, were without a known upper or lower extremity mobility problem, and who were hospitalized for a month or more after birth (N≈700;rounded to the nearest 50).

Directed Acyclic Graph

We constructed a Directed Acyclic Graph and enumerated all open door pathways (Figure 4). The following confounders were included in the minimally sufficient conditioning set to block all confounding pathways: gestational age, length of the child's hospital stay after birth, age of early motor milestone attainment, parental education, race/ethnicity, socioeconomic status, health condition, birth weight, weekly hours in childcare, stimulation in the home, 9 month BSF-R Motor T Score.

Modeling Variables Associated with Exposure and Outcome in the Unexposed:

We evaluated the association between covariates and receipt of therapy services as well as preschool motor performance in the unexposed. We chose the skipping item as a strong indicator of motor coordination. Children with a visual deficit (OR: 0.77 95% CI: 0.22, 2.70) or with a hospital stay after birth of more than a month (45-60 days; OR: 0.74 95% CI 0.33, 1.64) had a decreased odds of successfully skipping eight consecutive steps. Increased age at which the child cruised (0.83, 95% CI 0.66, 1.05) was also associated with a decreased odds of passing the skipping assessment. Children who spent more than 40 hours per week in childcare were 0.56 times (OR: 0.56 95% CI 0.23, 1.46) as likely to be successful with the skipping task compared to children who were in childcare less than ten hours per week. Children whose parents had a high school education or some college were 20% to 80% more likely to skip eight consecutive steps successfully compared to children without a high school education (Table 8).

In this sample, length of hospital stay and age at which the child achieved early developmental motor milestones were strongly associated with receipt of either occupational or physical therapy. (Table 8) Also, children who had a visual deficit were 4.54 times as likely (OR: 4.54, 95% CI: 2.44, 8.45) to receive services compared to children without a visual deficit. With every ten unit increase in the child's BSF-R Motor T Score, the odds of (OR: 0.55, 95% CI: 0.43, 0.69) receiving therapy decreased by 45%. Children, whose families provided greater stimulation in the home, were also 89% more likely to receive therapy, than families who appeared to provide an average level of stimulation in the home (Table 8). When compared to white non-Hispanic children, African American children were less likely to receive therapy (OR:0.53, 95% CI: 0.29,0.97) services, yet more likely to successfully (OR:3.18, 95% CI 1.75,5.78) complete the skipping task.

We also explored potential confounders by calculating the percentage change in the effect estimate with a given covariate dropped from the full model. The change is estimate was calculated as follows ln(OR full/ORreduced). The following covariates changed the effect estimate more than thirty percent when dropped from the full model: length of the child's hospital stay, parental education, Medicaid status, APGAR score, and age at cruising. The effect estimate also changed more than ten percent when race or birth weight was dropped from the full model (Table 9).

Based on these analyses of confounding, we included the following strong confounders in our propensity score model (Table 10): gestational age, birth weight, length of the child's hospital stay, cruising, race/ethnicity, parental education, socioeconomic status, and 9 month BSF-R Motor T score. Age at which the child cruised and length of time the child remained in the hospital after birth appeared to be the strongest confounders. We conducted a sensitivity analysis comparing effect estimates and the precision of our estimates including all potential confounders (Model 1) (strong and weak) compared to a propensity score model including only strong confounders (Model 2) (Table 11). The effect estimates from the more parsimonious propensity score models showed similar effect estimates, for most outcomes, with greater precision. Therefore, when considering the bias precision trade off, we chose Model A to estimate the predicted probability of therapy between nine months and age two.

Very Low Birth Weight Cohort:

At risk for delayed motor coordination in preschool

In our preliminary work, we examined the relationship between confounders and preschool motor ability in a population of low birth weight children with similar functional ability. Here, we further define a sample of very low birth weight children who are at risk for impairment in motor coordination in preschool. This sample included very low birth weight children, without known congenital anomalies, who appeared to be reaching normal developmental milestones (cruising on the nine month assessment) and did not have a known medical diagnosis that would affect mobility (Figure 6). Although the investigators obtained these indicators by parent-completed questionnaire, studies show that parents provide dependable reports of their child's motor ability and health.[83 84]

I. Crude Model

Crude models estimating the effect of interventional therapy services and preschool motor coordination showed reverse causality. Very low birth weight children who received the intervention were approximately 20%- 30% less likely to be successful with preschool motor coordination skills (Table 12).

II. Standard Outcome Model

We entered the following confounders in the model according to the functional form of the relationship depicted in our exploratory analysis (Figures 8-12; 21-26) : days in hospital (continuous) age cruising (continuous), 9 month BSF-R Motor T score (continuous), birth weight (continuous), gestational age (continuous), race (indicator variables), education (indicator variables), socioeconomic status (restricted quadratic spline; knots -1.5,0.3,1.2). Using logistic regression in SAS version 9.2, receipt of interventional therapy was associated with improved preschool motor coordination among typically developing very low birth weight children. Very low birth weight children who received the treatment were 1.67 times as likely to successfully skip eight consecutive steps (OR: 1.67; 95% CI: 0.46, 6.03) when compared to children who did not receive the treatment. There was no effect of therapy on walking backwards (OR: 1.03, 95% CI: 0.37, 2.88). Children who received the treatment appeared to do marginally worse on balance and hopping skills (Table 13).

III. Missing Data

Approximately 7% of children were missing at least one covariate used to estimate the propensity score. Therapy status was unknown for twenty seven percent of these children. Ninety seven percent of children with a missing covariate where missing either the *9 month BSF-R T score* or gestational age. The majority of children missing a covariate used to estimate the propensity score (67%) did not receive treatment. The distribution of strong confounding covariates among children not included in estimating the propensity score was similar to the untreated group in the full sample.

Approximately 20% of our VLBW sample was missing data on their ability to perform items measuring preschool motor coordination. Average birth weight (1148 grams vs. 1127 grams), length of hospital stay (49 days vs. 50.85 days), and 9 month BSF-R motor scores (49.63 vs. 48.76) were similar between children with preschool motor scores and those with missing data respectively. Families with missing information on preschool motor performance scored lower on the socioeconomic scale (-0.21 vs. -0.40) compared to families with data on these preschool endpoints.

Chapter 4

PHYSICAL AND OCCUPATIONAL THERAPY SERVICES AND MOTOR COORDINATION AMONG CHILDREN BORN VERY LOW BIRTH WEIGHT[c]

Introduction

Since the 1980's, rates of infants born with low birth weight (LBW) have increased in the United States. In 2010, approximately 8.2% of all births were LBW and 1.5% of all births were very low birth weight (VLBW).[1]

A small percentage of VLBW children (≈10%) suffer from severe neurological impairments including abnormalities in tone, transitional movement, and persistence of primitive reflexes. A large proportion experience only minor motor difficulties with complex movement.[2,3] Children born VLBW without notable neurological deficit show an initial delay in foundational motor skills. However, these children often catch up to children of normal birth weight during the first few years of life. As VLBW children approach elementary school age and motor skills become more complex, there is once again a divergence in motor ability.[33]

---

[c] Watkins, Stephanie MSPH, MSPT, Jonsson-Funk, M. PhD, Brookhart M A PhD, Rosenberg S A PhD O'Shea T M MD, MPH, Daniels J PhD. Physical and Occupational Therapy Services and Motor Coordination among Children Born Very Low Birth Weight..2012 (Under Review: Pediatrics).

Specifically, many VLBW children experience difficulty with motor coordination when they approach school age. Although they typically attain foundational motor milestones, these children are challenged to learn new motor tasks involving balance and coordination. They may appear “clumsy” or “awkward” and have difficulty with daily activities and classroom skills such as tying shoes or participating in physical education.[4]

In the absence of a neurological or medical diagnosis, these symptoms are described as developmental coordination disorder (DCD).[10] In a meta-analysis, the prevalence of DCD among VLBW children has been reported to be as high as 72%. Moreover, VLBW children are six times as likely to develop DCD than children born with normal birth weight.[10]

Poor motor coordination in childhood has a negative effect on the child’s mental and physical health. Many children with impaired coordination avoid social situations, as well as classroom activities and recreational activities placing them at risk for low self-esteem, social isolation, low levels of physical activity and obesity which persists into adolescence.[5,6-9]

To attenuate the potential negative sequelae of poor childhood motor coordination, early intervention by physical and occupational therapists is recommended.[10] Services are targeted toward facilitating motor control and motor planning that may improve motor coordination, overall levels of physical activity, and self-esteem at school age.

In the published literature, few studies have examined the efficacy of interventional physical and occupational therapy services on school age motor skills of low birth weight children. In a population of VLBW infants without neurological involvement, two small randomized studies assessed the effect of occupational therapy and physical therapy during the first year of life on motor ability in childhood.[11,12] Neither trial found a statistically significant difference in standardized motor scores between treatment groups. However, among children born of normal birth weight, interventions promoting motor development appear to improve children's locomotor ability.[13]

Little is known regarding the efficacy of physical and occupational therapy on childhood motor coordination. Therefore, in a population of VLBW children at risk for developmental coordination disorder, we estimated the effect of interventional physical and occupational therapy services delivered between nine months and two years of age on preschool motor coordination and change in motor skills between preschool and kindergarten.

Methods

Data Source

We used data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B), sponsored by the United States Department of Education and the National Center for Education Statistics. The ECLS-B is a nationally representative sample of children born in the United States in 2001 and followed through kindergarten.[85]

The ECLS-B collected data on children's physical, social, emotional, and cognitive development as well as health and education at four time points: nine months (2001-2002), two years (2003-2004), preschool (2005), and kindergarten (2006-2007). Specifically, at each wave of data collection, caregivers completed a self-administered questionnaire reporting demographic information, family structure, child health, family health, information on the home environment, parental attitudes, child development, and community support. Researchers directly assessed children's cognitive performance, socio-emotional development, and physical performance over each follow up period.[85]

Researchers oversampled selected demographic groups including: Asian, Pacific Islander and Chinese children as well as twins and children born low birth weight. Children who were born to mother's less than 15 years old, children who were adopted, and children who died before nine months of age were excluded from the cohort. At the nine month assessment, about 10,700[d] children and their parents participated in data collection.

Study Population

From participating families, we identified a sample of 500 VLBW children (<1500 grams) who were at risk for developing developmental coordination disorder. This sample included children who appeared to be developing normal motor skills in infancy and who did not have a known medical diagnosis that would affect mobility. Motor skills develop in typical sequential patterns which are

[d] All numbers are rounded to the nearest 50 to protect the confidentiality of participating families

described by developmental milestones.[86] On the nine month interview (mean age of assessment 11.9 months) parents reported the age at which their child attained the following motor milestones: sitting independently, crawling, pulling to stand, and walking with support. These motor milestones are fundamental to the development of upright locomotion and provide a framework for monitoring development over time.[87,88] We considered children who were successfully walking with support, the most advanced skill in this developmental sequence, as meeting normal developmental milestones.

In addition, on the nine month interview, families reported whether a physician identified their child as having difficulty with either upper or lower extremity mobility problems. We used this question as an indicator of children with a medical diagnosis that could potentially affect the child's motor ability. (Figure 6) We excluded children with congenital anomalies and known upper or lower extremity mobility impairments. Although the investigators obtained these indicators by parent-completed questionnaire, studies show that parents provide dependable reports of their child's motor ability and health.[83,84,89]

Measures

At the two year follow up assessment, all caregivers were asked if their child or family received services to help with their child's special needs. Families reported the type of service that the child received since the previous nine month assessment. We classified children as exposed to physical or occupational therapy services if the family indicated that the child ever received either type of service between nine months and two years of age.

The Bruininks-Oseretsky Test of Motor Proficiency (Bruininks) and the Movement Assessment Battery for Children (MABC) are norm referenced assessments used to identify children with mild to moderate impairments in motor ability among children between the ages of 4 to 21 and 3to17 respectively. The instruments are intended for use by practitioners to screen for motor impairments, support diagnoses of motor impairment, to assist with placement, and to evaluate the effect of interventions.[62,64,90]

The preschool assessment included direct measures of children's gross motor ability adapted from both the Bruininks and the MABC. Investigators did not administer the gross motor assessment from these measures in their entirety. Rather, ECLS-B selected items from these assessments emphasizing strength, agility, and motor coordination. Investigators asked participating children to complete each of the following items: skipping eight consecutive steps, hopping on one foot five times, walking backwards six steps on a taped line, standing on one foot for ten seconds, and jumping forward from a standing position.

Children received one trial to complete the skipping and walking backwards items and three trials to complete the balance and hopping items. Investigators scored the items on a pass/ fail basis. For the balance and hopping items, we classified the child as passing the item if they completed the skill on either foot. For the jumping task, each child received two trials, and investigators reported the greatest distance the child jumped in inches.

The assessment was repeated at the kindergarten follow up assessment. Thus, we calculated the change in jumping distance between preschool and kindergarten. The other items were scored as present or absent and did not provide a continuous scale for us to assess change in the skill level between preschool and kindergarten.

Confounders

We considered a list of covariates, measured on the 9 month questionnaire, as potential confounders of the relation between receiving therapy services in the toddler years (9 months to 2 years) and motor performance in preschool.[e] To determine a final list of confounders, we considered both *a priori* substantive knowledge and analyses of confounding in the data set. We created a directed acyclic graph to identify a minimally sufficient conditioning set of variables to control confounding.[65] We also used logistic regression to model the relation between covariates and therapy receipt as well as the relation between covariates and preschool skipping ability among children who did not receive the treatment. The final covariates included in the analysis were variables in the minimally sufficient conditioning set and those variables associated with only the outcome.[66] The covariates included: gestational age, birth weight, length of the child's hospital stay after birth, age at which the child began to walking while

[e] Maternal depression (Center for Epidemiologic Studies-Depression Scale), child stimulation in the home environment (Home Observation for Measurement of the Environment- Short Form), parental knowledge of infant behavioral norms and child developmental processes (Knowledge of Infant Development Inventory), birth weight, number of siblings, gestational age, 5 minute APGAR scores, length of hospital stay after birth, length of NICU stay after birth, 9 month BSF-R motor T score, socioeconomic status, hours per week in childcare, age of early motor milestone attainment, parental education, health condition of the caregiver, history of childhood injury, health condition of the child, single parent status, social support with childcare, health insurance status, race/ethnicity, maternal work schedule, urbanicity, and geographic region

holding onto furniture, race/ethnicity, parental education, socioeconomic status, and the child's 9 month BSF-R T score.

Statistical Analysis

In this non-experimental study, treatment assignment into physical or occupational therapy is not random. Children who received physical and occupational therapy services may differ from the children who did not receive therapy on demographic characteristics and baseline severity of their functional ability. In particular, children who received therapy may be "more functionally limited" than the untreated children resulting in confounding by severity of the child's health.

To estimate the average treatment effect of early childhood physical and occupational therapy on preschool motor skills, we used propensity score methods to control for confounding. The propensity score is the conditional probability of receiving treatment given a set of observed covariates. Once we condition on the propensity score, receipt of therapy should then be independent of covariate patterns, assuming no unmeasured confounders.[14]

We estimated the conditional probability of treatment given our covariates using a Random Forest model.[69] This approach is based on an aggregation of classification trees, each built using a recursive partitioning algorithm. Researchers have shown in simulation studies that propensity scores estimated using machine learning methods, such as the Random Forest algorithm, may reduce bias when compared to standard logistic regression.[17,18]

We therefore fit the Random Forest with 1000 trees using 4 variables chosen randomly at each split, given the covariates listed above using the Random Forest package in R version 2.6.1.[91] The analysis included all confounders and variables associated with the outcome in the algorithm.[66] All children with complete data on confounders used to estimate the propensity score were included in the Random Forest model (93.04% of sample).

We used the propensity score to create inverse probability of treatment weights. Children who received the treatment received a weight of (1/propensity score) and children who did not receive treatment received a weight of (1/ (1-propensity score)). By weighting each individual in this manner, we create a "pseudo population" in which the distributions of covariates in each of the two treatment groups should mirror the covariate distributions of the original combined sample. To assess the comparability of the distribution of confounding covariates after weighting the sample, we calculated the standardized difference of confounding covariates between the treated and untreated children.[f] If the covariates are similarly distributed after weighting, then the treatment group is no longer associated with risk factors for the outcome, and the comparison of

[f] Standardized differences represent the differences between the means by therapy status in units of standard deviations. The estimates is calculated as $d=(xbar_{therapy}-xbar_{no\ therapy})/sqrt(S^2_{therapy} + S^2_{no\ therapy})/2))$75.Flury BK, Riedwyl H. Standard distance in univariate and multivariate analysis. The American Statistician 1986;40:249-51.

outcomes in the two groups should be an unbiased estimate of the average treatment effect of early childhood therapy.[92]

We then modeled the association between receipt of physical and occupational therapy between nine months and age two and the odds of performing select motor tasks in preschool in SAS version 9.2. Using logistic regression, we generated separate multiple regression models to estimate the average effect of physical and occupational therapy on the ability of the child to pass preschool motor outcomes. In addition, we modeled the effect of therapy on the distance that the child could jump using linear regression. To limit the potential for residual confounding, we adjusted both logistic and linear regression models for strong confounding variables including: birth weight, length of the child's hospital stay after birth, and age at which the child began to walk with assistance, all included as continuous terms into the model.

Finally, we calculated the change in jumping distance between preschool and kindergarten. We used linear regression to model the average change in jumping distance by children's exposure to early childhood physical and occupational therapy services. We adjusted this model for the child's jumping distance at baseline (preschool) and age when the child first began to walk with assistance.

Results

Descriptive Statistics

Our sample included about 500 VLBW children who were at risk for impairments in motor coordination during early childhood. Six percent of children in the sample received physical or occupational therapy services between nine months and two years of age. Treatment status was unknown for ten percent of the sample and these children were not included in the analysis. Children who received therapy were more likely to be male (61% vs. 45%), and white (58% vs. 38%). The mean birth weight for children receiving treatment was 1,030 grams (STD: 212.63) compared to the untreated children with a mean birth weight of 1,148 grams (STD: 250.31). Children in the therapy group had a longer mean hospital stay after birth (68 days vs. 48 days) and were born on average two weeks earlier than children in the non-treated group (28 weeks vs. 30 weeks). Developmentally, the treated children sat independently and crawled one month later, on average, than the untreated children. Five minute APGAR scores were similar between the two groups (Table 13).

Missing Data

Children without complete data on the covariates used to estimate the propensity score were excluded from the analysis. Approximately 7% of children were missing at least one covariate used to estimate the propensity score. Therapy status was unknown for twenty seven percent of these children. Ninety seven percent of children with a missing covariate where missing either the 9 month BSF-R T score or gestational age. The majority of children missing a covariate used to estimate the propensity score (67%) did not receive treatment. The distribution of strong confounding covariates among children not included in estimating the propensity score was similar to the untreated group in the full sample.

Approximately 20% of our VLBW sample were missing data on their ability to perform items measuring preschool motor coordination. Average birth weight (1148 grams vs. 1127 grams), length of hospital stay (49 days vs. 50.85 days), and 9 month BSF-R motor scores (49.63 vs. 48.76) were similar between children with preschool motor scores and those with missing data respectively. Families with missing information on preschool motor performance scored lower on the socioeconomic scale (-0.21 vs. -0.40) compared to families with data on these preschool endpoints.

Covariate Balance

A standardized difference, representing the difference in the means of a covariate by treatment status, of less than 0.10 is suggested as a cut point indicating only a minimal difference between treatment groups.[76,77] In the weighted sample, the distribution of baseline covariates was similar (<0.10) by receipt of physical and occupational therapy. The standardized difference between age of length of hospital stay (0.14), gestational age (0.15), and birth weight (0.08) were marginally different by therapy status (Table 14). In the weighted sample, children who received therapy stayed in the hospital on average ten days longer, were born one week earlier, and weighed 44 grams less than the children who did not receive therapy. However, the means standardized difference across all covariates was 0.07. Overall the covariates were well balanced between treatment groups.

Multivariable Regression

In the weighted sample, receipt of physical or occupational therapy between nine months and age two was strongly associated with the ability to perform coordinated movement in preschool. Specifically, children who received physical or occupational therapy between nine months and two years of age were 2.39 times (95% CI: 0.76, 7.51) as likely to skip eight consecutive steps compared to children who did not receive these services during early childhood, though the estimate was imprecise. Children who received therapy were also somewhat more likely (OR: 1.52, 95% CI: 0.51, 4.54) to successfully walk backwards six steps consecutively on a line when compared to children who did not receive the treatment, though this estimate was also imprecise. There was no association between therapy receipt and balancing independently on one leg (OR: 1.07, 95% CI: 0.45, 3.13 nor successfully hopping on one foot independently (OR: 0.90, 95% CI: 0.33, 2.45).

In the linear regression analysis, children who received therapy jumped slightly further than children who did not receive therapy, although not statistically significant at conventional levels. With regard to jumping performance between preschool and kindergarten, therapy appeared to have little effect on the change in jumping distance during this time period (Table 15).

Sensitivity Analysis

We examined the sensitivity of our effect estimates by restricting our sample to those children with overlapping propensity scores. Twenty five percent of the sample was excluded. From our sample with overlapping propensity scores, we also excluded children who were treated most contrary to prediction. For example, we excluded children who received therapy but had a propensity for treatment lower than the 1st percentile and children who did not receive therapy but had a propensity for treatment greater than the 99th percentile. We also trimmed the sample using the 2.5th and 97.5th percentile cut points for the treated and untreated children respectively.[80] Together, these trimmed percentiles excluded an additional 1% of the sample with overlapping propensity scores. Simulation studies show that asymmetric trimming of the propensity score leads to a reduction in bias in the estimate of a uniform effect of treatment in the presence of unmeasured confounding.[80] Effect estimates were similar across the unrestricted sample, the sample with overlapping propensity scores, and across the trimmed samples (Table 16).

Discussion

In a sample of VLBW children who were "at risk" for developmental coordination disorder, receipt of physical or occupational therapy services between nine months and two years of age was associated with the child's ability to perform certain coordinated movement tasks in preschool. Although not statistically significant, when compared to the untreated children, therapy may also improve lower extremity strength as suggested by the child's jumping distance. Although our estimates were imprecise, these findings suggest physical and occupational therapy services delivered during the toddler years may improve a child's functional ability to perform refined coordinated movement in early childhood.

Children who are born VLBW are at an increased risk for developmental coordination disorder in childhood.[10] Although initially many of these children attain foundational developmental milestones, motor skills that involve balance and coordination become challenging. Children who experience difficulty with motor coordination frequently feel less confident around their peers and may withdraw from social situations involving physical activity.[7] This avoidance behavior may set a precedent for low levels of fitness since proficiency of motor skills during this early childhood developmental period is an important predictor of physical activity into adolescence.[93] Improvement in school age motor coordination may promote the child's confidence to engage in leisure time physical activity with peers and to participate in organized sports improving their overall mental and physical health.

The literature evaluating the efficacy of occupational and physical therapy services to improve motor coordination among low birth weight infants is sparse. In a meta-analysis of preschool age children of normal birth weight, interventions that promoted motor skills appeared to improve early childhood object control and locomotor skills.[13] However, two small randomized studies evaluating the effect of physical therapy and occupational therapy, delivered before twelve months of age did not find a statistically significant difference in school age motor scores compared to controls.[11,12] These studies evaluated therapy services in a small population (< 100 children) of VLBW children without neurological problems.

Previous studies evaluated the effect of physical and occupational therapy delivered during the child's first year of life. The focus of treatment during this developmental window was most likely on fluidity of movement and independent transitions with pre ambulatory skills including head control, independent sitting, crawling, and walking with support. However, our work focuses on therapy between one and two years of age where the focus of treatment may include facilitation of foundational skills of ambulation, strengthening, coordination, and balance. This type of treatment may more directly carry over into improvement in more complex movement patterns.

Several limitations in our work should be noted. First, the ECLS-B data, while rich, only provided a crude measure of therapy receipt over an approximate fifteen month period. We were unable to assess whether children consistently received therapy over this time period or how their dose of therapy may have affected the results. Since therapy treatment plans in early intervention are frequently written for a year, there was most likely minimal transition out of services during this 15 month period. Yet, with only a crude measure of treatment, the magnitude of our results may be attenuated where accounting for degree of treatment may reveal more profound results. Additional studies are needed that provide information regarding the amount, frequency, and length of treatment to determine the optimal duration and dosing of therapy that will result in improved school age motor coordination in this population.

These data had rich measures of the child's demographic status, health, and functional ability that we accounted for in our analysis. Yet, although we used robust methods to account for self-selection into therapy, there is the possibility that residual confounding could have influenced our effect estimates. However, when we trimmed children from the analysis who were treated contrary to their probability of receiving therapy, our results were similar to our original estimates.

We also estimated the association between receipt of physical and occupational therapy services and preschool age motor coordination using only observations with complete data. Although the percent of children who were missing baseline covariates was small (<10%), approximately 20% of children did not have measures of motor skills in preschool. Therefore, due to missing data, our sample size was reduced which decreased the precision of our estimates. When compared to children without missing data, children with incomplete data were of similar birth weight, had a similar hospital stay after birth, and met developmental milestones on a similar time frame. However, the complete cases had a higher mean score on the socioeconomic scale. Thus, our results may not be fully generalizable to very low birth weight children from a lower socioeconomic status.

Nonetheless, our study has many strengths. Previous studies evaluated the effect of physical and occupational therapy delivered during the child's first year of life. The focus of treatment during this developmental window was most likely on fluidity of movement and independent transitions with pre ambulatory skills including head control, independent sitting, crawling, and cruising. However, our work focuses on therapy between one and two years of age where the focus of treatment may include facilitation of foundational skills of ambulation, strengthening, coordination, and balance. This type of treatment may more directly carry over into improvement in more complex movement patterns.

Moreover, our work evaluated acquisition of individual preschool motor skills compared to scores on a standardized assessment. Although standardized assessments are useful to compare children to a developmental norm, standard scores may be highly variable in infants and toddlers and the sensitivity of standard scores to detect motor coordination difficulties, especially among children with DCD, has been questioned in the literature.[94-97] We found that interventional therapy was beneficial for selective skills, where therapy appeared to have little effect on the achievement of other motor tasks. Assessment of individual skills may therefore be useful to evaluate effects of interventions in this population.

We used data from a prospective longitudinal study which was representative of children in the United States born in the year 2001. Researchers interviewed families at multiple developmental stages (9 months, 2 years, preschool age) which most likely limited poor recall of whether or not the child was receiving occupational or physical therapy services. Moreover, researchers implemented quality control procedures for direct child assessments. Even though the researchers did not assess reliability of the gross motor assessment, the field investigators agreed on the scoring of direct assessment of fine motor talks approximately 90% of the time.[85]

In our analysis, we used rigorous methods to both select appropriate confounding variables and to balance the distribution of these variables between treatment groups. We considered the sensitivity of our findings to children that did not have a similar need for treatment and who were treated contrary to their estimated propensity for treatment. Our findings were robust to these sensitivity analyses.

In a sample of children born VLBW who were at risk for poor motor coordination in childhood, we found that those who received interventional physical and occupational therapy between nine months and age two were more likely to successfully perform higher level coordination tasks compared to children who did not receive these services. Providing treatment to these children while ambulation, coordination, and balance skills are emerging may optimize health in function in school age children. The implications for improvement in motor performance may have far reaching effects on physical activity, school performance, and self-esteem perhaps into early adulthood.[98-100]

Chapter 5

AN EMPIRICAL COMPARISON OF TREE-BASED METHODS FOR PROPENSITY SCORE ESTIMATION: PHYSICAL AND OCCUPATIONAL THERAPY SERVICES AND PRESCHOOL AGE MOTOR ABILITY[g]

Introduction

Research treatment effectiveness in non-randomized studies is complicated by exposure group differences on measured and unmeasured characteristics that are independently related to the outcome of interest. Propensity scores are commonly used to control for confounding when estimating treatment effects in non-randomized studies. The propensity score is the probability of receiving treatment given confounders.[14] Subjects with similar propensity scores can be expected to have similar values on measured background characteristics. Once one conditions on the propensity score, differences in measured characteristics between the treatment groups should be from chance alone.[14]

[g] Stephanie Watkins MSPH, MSPT, Brookhart M A PhD, Jonsson-Funk M PhD, Rosenberg S. A. PhD, O'Shea M. MD MPH, Daniels J PhD. An empirical comparison of tree based methods for propensity score estimation: physical and occupational therapy services and preschool age motor ability.2012. (Under Review: Child Development).

Logistic regression is frequently used to estimate propensity scores, but requires several assumptions. The relation between continuous and ordinal independent variables and the dependent variable logit must be linear. Furthermore the joint effect between independent variables in the model must be considered, as well as the functional form of covariates or interaction terms.[15] Violations can result in misspecification of the propensity score model and the resulting effect estimate may be biased.[16]

Regression tree based methods, including Bagging and Random Forest classification (RFC), are non-parametric methods derived from learning based algorithms which offer alternative strategies for generating predicted probabilities of treatment. The methods use a series of classification trees to estimate the average probability of membership in a given class. These techniques have been suggested to have improved predictive accuracy when compared to classical statistical techniques. [59] For example, in simulation studies, regardless of non-linearity or non additivity, random forest performed well in terms of covariate balance between treatment groups and may result in further reduction in bias of the effect estimate when compared to traditional logistic regression.[17,18]

There has been relatively little investigation into the use of tree based methods to estimate the propensity score.[58] In this article, we illustrate the use of two tree based methods, bagging and RFC, in the context of an analysis to understand the effect physical and occupational therapy services on the motor skills of preschoolers who were born with very low birth weight (VLBW). We consider the propensity scores generated by bagging and RFC as well as two additional estimation approaches: a single classification tree and a main effects logistic regression model. We then compare the distribution of the estimated propensity scores and both the balance of covariates as well as effect estimates after applying inverse probability of treatment weights (IPTW).

Conceptual Overview

Classification Trees

Classification tree analysis is a non-parametric method commonly used in data mining where a set of independent variables are used to predict membership of observations in a given class of the dependent variable. The method evaluates the relationship between predictors and treatment with a learning algorithm using decision trees to partition observations into nodes with similar probabilities of class membership in the treatment group.[59] The data set is partitioned until nodes, or branches of the tree, are as homogenous as possible with respect to class membership.[101] The tree begins with a root node and continues to split until the nodes reach either a given sample size or a given level of impurity reduction. At each terminal node, the algorithm predicts the response class by taking the majority vote from all of the observations within a given node.[68]

Despite the popularity of this data mining method, results from a single classification tree are highly variable and are known to be highly unstable. For example, the rank of each variable in the classification tree as well as the cut point of the variable is strongly dependent upon the distribution of observations in the data. With small changes in the data structure, the order of variable selection or the cut point of the variable may change resulting in an alternative tree structure.[68]

Bagging and Random Forest Classification

Both bagging and RFC are tree based methods that attempt to improve the stability of tree based regression methods based on single trees. These methods aggregate predictions over multiple individual classification trees to improve the overall predictive performance of the algorithm. Bagging randomly draws a series of bootstrap samples from the data, and creates individual classification trees for each sample. With each of these bootstrap samples, the data will vary slightly from the previous sample. Furthermore, each individual tree may then vary, perhaps substantially, from the previous tree. The algorithm then aggregates the predicted probability of class membership over the series of classification trees.[69]

Random forest classification utilizes this same bootstrap method. However, random forest adds an additional level of variability to the algorithm. During construction of the individual classification trees, a random sample of predictor variables is chosen to split the data at each node. Therefore, each individual tree is even more diverse compared to the trees from bagging alone.[68]

Although individual classification trees are inherently unstable, bagging and random forest classification have been shown to produce robust estimates. In both empirical and simulation studies, estimates aggregated over a series of classification trees show improvements in prediction accuracy when compared to a single classification tree.[70-73] Bagging is suggested to equalize the influence of given observations in the data. Thus, data points which strongly influence the classification algorithm are downweighted.[68] Furthermore, the additional level of randomness introduced by RFC creates additional diversity between trees with a lower upper bound of error.[69] Overall, these methods produce a more robust final estimate with decreased variability.[69]

Methods

We illustrate the use of three tree based methods: bagging , RFC , and a single classification tree, as well as parametric logistic regression in an analysis that evaluates the effect of physical and occupational therapy services on motor performance among preschool children who are VLBW and “at risk” for developmental coordination disorder (DCD). DCD is a condition defined as impairment in the development of motor coordination among children without known physical or neurological impairments.[36] Children with DCD are at an increased risk for low academic performance, low self-esteem, and limited physical activity which may continue into adolescence. Children who are born with VLBW are six times as likely to have DCD compared to their normal birth weight peers.[10]

Population and Variables

Using data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B), our sample included approximately 500 VLBW children who were without known mobility problems and appeared to be meeting normal developmental motor milestones at nine months. Researchers asked families between nine months and age two if their child had ever received physical or occupational therapy services. We considered the child exposed if the child ever received either therapy during this time period.

Researchers directly assessed preschool gross motor performance using items from the Bruininks-Oseretsky Test of Motor Proficiency (Bruininks) and the Movement Assessment Battery for Children (MABC). Both assessments are norm referenced and designed to assess both gross and fine motor functioning from childhood through adolescence.[62] Researchers directly reported the child's ability to complete the following tasks on a pass fail basis: skipping eight consecutive steps, hopping on one foot five times, walking backwards six steps on a taped line, and standing on one foot for ten seconds.

Based on a priori substantive knowledge, we created a directed acyclic graph and determined a minimum sufficient conditioning set of confounders. The final covariates in our analysis included: gestational age, birth weight, length of the child's hospital stay after birth, age at which the child began to walk with assistance, race/ethnicity, parental education, socioeconomic status, and the child's nine month Bayley Short Form-Research Edition (BSF-R) motor T score.

The BSF-R is a subset of items taken from the standardized Bayley Scales of Infant Development Second Edition to assess children's cognitive, motor, and language skills.

Propensity Score

In this observational study, treatment assignment into physical and occupational therapy services was not randomized. Children and their families may participate in therapy treatment based on a host of factors including their child's functional ability and access to healthcare. Therefore, the distribution of baseline characteristics between children in the treated and untreated groups may differ and children between these two groups would not be "exchangeable". We estimated the average treatment effect of early childhood physical and occupational therapy using a propensity score approach to control for confounding.

Estimating the Propensity Score

We estimated the conditional probability of treatment given the identified confounders stated above using the following four methods: bagging, random forest classification, a single classification tree, and logistic regression.

Using the R statistical platform we first used the RandomForest package to estimate the predicted probability of class membership in the therapy group given the following covariates: gestational age, birth weight, length of the child's hospital stay after birth, age at which the child began to walk with assistance, race/ethnicity, parental education, socioeconomic status, and the child's nine month BSF-R Motor T score. Race/ethnicity and parental education were entered as a series of indicator variables; all other variables were entered as continuous variables. All categorical variables were encoded as "factors" in the R environment. This transformation ensured that the R software recognized these variables as categorical responses.

We set the random forest algorithm to generate 1,000 individual classification trees. The suggested default for the number of random splitting variables at each node is the square root of the number of variables in the algorithm. Our model included 19 variables, so we set the default to 4 variables chosen at each split.

We checked the sensitivity of the error rate to our chosen parameters by allowing the number of trees to vary between 250 and 1,000 and the number of randomly chosen variables to vary between 2 to 7. The error rate for the algorithm is generated from the 33% of the data remaining that was not used to form the classification trees. For example, with each bootstrap sample, the remaining data (≈33%) not in the sample is entered into the classification tree. The error in these out of bag predictions is collected over the series of trees to determine the final error rate over the forest. The error rate is suggested to be robust if the predicted probabilities of class membership are aggregated across a sufficient number of trees. However, if the number of trees are too few, then the error rate may be upwardly biased.[74] The algorithm may therefore be a better predictor of the outcome than suggested by the error rate.

We then implemented the Ipred package and the Tree package using the R statistical software to estimate the predicted probabilities of having class membership in the treatment group using bagging and a single classification tree respectively. For both models, we entered the same covariates as in the RFC algorithm. In the Ipred package, we generated a series of 1,000 trees and checked the sensitivity of the error rate by varying the number of trees between 250 and 1,000. For both methods, the splitting variables were chosen by the algorithm in a hierarchical fashion based on impurity reduction.

Lastly, we generated predicted probabilities of receiving physical or occupational therapy using logistic regression. As in common practice, we entered potential confounders as main effects. Race/ethnicity and parental education were modeled as indicator variables; all others were entered into the model as continuous terms.

Statistical Analysis

We generated unique inverse probability of treatment weights using each method: RFC, bagging, a single classification tree, and logistic regression. These weights create a pseudo population of children with a distribution of covariates that represents the combined sample.[92] To estimate the average treatment effect, treated children received a weight of (1/propensity score).

Children in the untreated group received a weight of (1/ (1-propensity score)). To evaluate the balance of each propensity score method, we then calculated the standardized difference of the weighted confounding variables between the treatment groups.

Standardized differences represent the differences between the means by therapy status in units of standard deviations. The estimate is calculated as $d=(xbar_{therapy}-xbar_{no\ therapy})/sqrt(S^2_{therapy} + S^2_{no\ therapy})/2))$[75]. Although there is no standard criterion to determine balance between treatment groups, experts suggest a standardized difference of <0.10.[76-78] We then averaged the standardized differences across all confounders to determine the mean standardized difference (MSD).

Finally, for each of the four methods, we estimated the average effect of physical and occupational therapy on preschool motor performance using logistic regression and inverse probability of treatment weights (IPTW) in SAS version 9.2. We controlled for birth weight, age at which the child walked with assistance, and the child's length of hospital stay after birth. The relationship between length of the child's hospital stay, birth weight, and age at which the child walked with assistance and the log odds of preschool motor development appeared linear and were entered as continuous variables. We stabilized the weights to obtain a narrower confidence interval around the estimated effect estimate by multiplying the child's IPTW by the probability of receiving the treatment that they actually received.[102]

Missing Data

In these data, approximately 7% of children were missing at least one covariate used to estimate the propensity score. We therefore included only children with complete data to estimate the predicted probability of treatment. Thus, we compare the balance of covariates and the estimated effect estimates for each method among the same group of children.

Results

The sample included approximately 500[h] children weighing less than 1500 grams at birth of which 6.5% of children received therapy between nine months and age two. Children who received therapy were more likely to be white (58.06% vs. 38.48%) and male (61.29 vs. 45.32%) and were born on average two weeks earlier in gestation. Developmentally, the treated children sat independently, crawled, and walked with assistance on average, one month later than the untreated children. Five minute APGAR scores were similar between the two groups (Table 13).

Random Forest Classification/Bagging: Error Rate

In our sample of approximately 450 children with complete covariate data, the algorithm misclassified treatment status 15.65% of the time over 1,000 trees with 4 variables randomly chosen at each split. Overall, there was little change in

[h] Numbers are rounded to the nearest 50 for data security

the error rate with small changes in the number of splitting variables. The error rate over our chosen range of trees and number of splitting variables varied by approximately 0.50%. The misclassification rate for the bagging algorithm over 1000 trees was 15.65%. The misclassification rate increased to 17.10% with only 250 trees (Table 17).

Propensity Score

The mean predicted probability of receiving treatment for the children who received therapy ranged between 0.16 to 0.20 across the RFC and bagging tree based methods and the main effects logistic regression model. The single classification tree yielded a predicted probability of treatment that was approximately twice that of the other three methods for children who received therapy. The mean predicted probability of treatment for children who did not receive therapy ranged between 0.05 and 0.07 across all four methods used to generate the propensity score. Children in the treatment groups received similar weights across estimation methods with the exception of the single classification tree algorithm. Children had a higher propensity for treatment and received a lower weight compared to the other estimation methods. The weights for children who did not receive physical or occupational therapy were similar for all four methods (Table 18).

Covariate Balance

In the unweighted sample, the MSD across strong confounding covariates was 0.54. The length of the infant's hospital stay after birth (Standardized Difference: 0.73) and the age at which the child crawled and walked with assistance (Standardized Difference: 0.74 and 0.64 respectively) had the greatest imbalance between the treatment groups. After applying the weights generated by the RF and bagging tree based methods, only a negligible difference existed in the distribution of baseline covariates by treatment status. The MSD across covariates was 0.07 using the random forest method and 0.03 using the bagging algorithm. After implementing the random forest algorithm, length of hospital stay and birth weight remained slightly unbalanced (Standardized Difference: 0.14 and 0.08 respectively). The mean length of hospital stay and birth weight after applying the bagging method was quite similar (Standardized difference: 0.03 and 0.02 respectively) by therapy status (Table 19).

The MSD for the covariates weighted with the logistic model was 0.11. The standardized difference for birth weight, length of hospital stay, and age at crawling and walking with assistance was greater than the suggested 0.10 criterion for these covariates. The propensity score estimated by the single classification tree demonstrated a poor ability to balance the covariates between treatment groups. These covariates continued to differ by approximately 0.18 standard deviations (Table 19).

Multivariable Regression

Overall, in the weighted multivariable logistic regression models, receipt of interventional physical or occupational therapy services between nine months and age two was associated with improvement in preschool coordination skills in this VLBW population. The effect was consistent across both the tree based methods as well as the logistic method used to generate the propensity score; however the magnitude of the effect as well as the precision of the estimate varied by method. The random forest algorithm produced the most precise estimate in the weighted model for hopping and single leg stance. The bagging algorithm produced slightly more precise estimates for the other preschool skills (Table 5). When we used logistic regression to estimate the propensity score, the confidence intervals for the effect estimates were the least precise. The magnitude of the estimate for skipping ability was largest (OR: 3.34, 95% CI 1.07, 10.44) using the bagging technique and smallest (OR: 2.35 95% CI: 0.71, 7.71) using logistic regression to estimate the propensity score. The bagging estimate continued to generate the effect estimates of the greatest magnitude for the additional motor outcomes modeled in these data. In general, logistic regression estimation of the propensity score produced the most conservative effect estimates for the majority of preschool motor items. The single classification tree algorithm did not balance the covariates well between treatment groups, and therefore the results of the weighted models using this method are not presented (Table 20).

Discussion

In this paper, we illustrated the use of various tree based methods to estimate the predicted probability of receiving interventional physical and occupational therapy services in a sample of VLBW children. Furthermore, we considered how propensity scores estimated from bagging and RFC balanced covariates between treatment groups and compared these methods with the performance of propensity scores estimated from a single classification tree as well as traditional logistic regression.

In our sample, bagging and RFC achieved the best overall balance of covariates across treatment groups. Among all methods used to estimate the propensity score, the mean standardized difference of all covariates was smallest for these two methods. The propensity scores estimated from the logistic model showed a marginal imbalance in covariates, where the single classification tree method had the worst performance.

These findings are supported by the study of Lee and colleagues who studied machine learning methods when estimating the propensity score in simulated data. In a small sample, when compared to standard logistic regression and a single classification tree, random forest and bagging returned the lowest mean absolute standardized differences. The standardized differences between individual covariates were also less dispersed with these two methods. The resulting bias in these simulated models was highest when the propensity score was estimated from a single classification tree and lowest when the propensity score was estimated using either bagging (10.3%) or RFC (7.7%).[17]

In our data, it appeared that propensity score estimation using logistic regression did a reasonable job of balancing the covariates between our treatment groups. However, it is unknown how well this model performed in reducing the amount of bias since the true treatment effect is unknown. While the effect estimates assessing preschool coordination were similar between the two models when we estimated the propensity score by RFC and logistic regression, the effect estimates for the child's ability to balance differed by approximately 13%.

In simulation studies, a main affects logistic regression model performed adequately in reducing bias when the relation between independent variables is linear and additive.[17] However, researchers reported a mean absolute bias of 30% in the presence of non additivity and non-linearity.[17] For comparison, we used a main effects only logistic regression model which appears to be commonly used by researchers. However, in our data, the relation between several confounders and the logit of receiving treatment was curvilinear. Due to our small sample size, we were limited in our ability to test for interactions. Therefore the difference in the estimated effects may be due to lack of consideration of the relation between confounders and the logit of receiving treatment. However, by modeling the functional form of the variable, for example including spline terms, and considering interactions the logistic regression model may be more effective. However the non-parametric random forest algorithm naturally incorporates interactions as well as non- linear functional forms which may be more feasible for the naïve researcher.

In our analysis, ensemble tree based methods, including random forest and bagging, appear to outperform traditional logistic regression methods with main effects. Both tree based methods performed well in balancing the covariates between treatment groups, however the bagging method resulted in effect estimates of greater magnitude. It is possible that the additional level of randomness implemented by the random forest classifier allowed less important variables to be expressed in predicting therapy exposure thereby attenuating the magnitude of the effects.

In addition to the improved performance of these methods over logistic regression, random forest performs well against other classifiers in simulated data.[69] For example, the predictive accuracy of a single classification tree is highly variable to small changes in the data structure. However, the variability introduced by aggregating responses over bootstrap samples of classification trees with variable tree structures improves the predictive accuracy of the algorithm.[68] Thus we would expect the predictive accuracy of the RF algorithm to also be robust to other populations of VLBW children with a similar distribution of baseline characteristics.

In this study, estimation of the propensity score using ensemble tree based methods produced the smallest standardized differences across covariates. The resulting effect estimates varied slightly depending on the method used to estimate the propensity score. Although, we are unsure of the true effect estimate, studies show that the effect estimates estimated from RFC and Bagging are the least biased and logistic regression may adequately reduce bias in the presence of non-additivity and non-linearity. However, in many epidemiological studies, many exposure outcome relations are complex in nature. Estimation of the propensity score using tree based ensemble methods may be a useful method to evaluate the effect of interventions on childhood motor skills. These methods appear to be a robust creating better covariate balance for control of confounding and further bias reduction compared to logistic regression.

Chapter 6

DISCUSSION

In this dissertation work, we estimated the effect of interventional therapy services delivered between nine months and two years of age on preschool motor skills in a population of low birth weight children. Interventional physical and occupational therapy services are frequently delivered to low birth weight children in early childhood to facilitate transitional movement and to promote strength, balance, and coordination.

It appears based on the published literature that therapy may have some beneficial effects on neurodevelopment, particularly among very low birth weight children.[46,51] The majority of present studies, specific to interventional therapy, evaluated treatment during the child's first year of life with follow up through age two. Yet, little is known regarding the efficacy of services on more complex motor ability later in childhood.

In these data, we first defined a population of children who weighed less than 2500 grams. This cohort comprised a heterogeneous group of low birth weight infants with diverse functional ability. Children in the treatment group had a lower mean baseline motor score, a longer average hospital stay after birth, and were significantly delayed on fundamental motor milestones compared to the children who didn't receive treatment. For example, forty six percent of low birth weight children in the treatment group were unable to sit independently without support during the nine month assessment, yet only 7% of children in the untreated group had not attained this motor milestone.

Due to such a disparate need for therapy between the two groups, we further defined a cohort of very low birth weight children with similar functional ability that were at risk for developing delays in motor coordination in preschool. These children appeared to be achieving early motor developmental milestones during an age appropriate developmental window and, at approximately 11 months of age (mean age of the nine month assessment), did not have a physician documented upper or lower extremity movement problem. Children who have difficulty with coordination often avoid participation in social play and physical activities that involve these skills. In turn, many children with impaired coordination have both low self-esteem and poor fitness levels.[5,7] Interventions which promote motor coordination among very low birth weight children are important since poor fitness levels in early childhood are important predictors of physical activity into adolescence.[93] Specifically, children who are born with very low birth weight are six times as likely to experience motor coordination disorder compared to children born with normal birth weight.[10]

In this sample of children with similar need for therapy, we then estimated the effect of interventional physical and occupational therapy on motor coordination in preschool. Very low birth weight children who received either physical or occupational therapy between nine months and two years of age showed improved ability in preschool motor coordination compared to children who did not receive services during this time period, although the estimate was imprecise. Few studies have evaluated the effect of therapy services on child motor development after age two in this population. Yet, these results support those of Riethmuller and colleagues who, in a systematic review, determined that interventions promoting motor development among children of normal birth weight, improved children's locomotor ability at preschool/school age.[13] Specifically, in our study, therapy was beneficial for a child's ability to walk backwards on a line and to skip eight consecutive steps.

Despite these encouraging findings, there appeared to be no effect of therapy on the child's ability to balance, while children in the treatment group, although not statistically significant, appeared to do marginally worse with hopping. Both hopping and maintaining single leg stance involve components of both strength and motor coordination. There is the possibility that therapy may improve performance with these skills, yet with greater intensity or perhaps duration of treatment that was not reflected in our crude measure of treatment.

ECLS B captured whether a child received treatment for individual therapies between nine months and two years of age. Yet, the amount of therapy that they received was reported for all therapies, including special education, during that time period. We were unable to accurately estimate the amount of treatment per month from an estimate that included multiple disciplines. There is a strong possibility that there may be a dose response relationship between intensity of physical and occupational therapy and success with preschool motor skills. For example, a greater amount of therapy during the toddler years may be required for a child to achieve success with motor skills that involve components of strength and motor coordination compared to skipping, which lacks the strength component. With additional information regarding the child's intensity of treatment, our effect estimates may have been larger in magnitude. By using such a crude estimate, with overall regression of intensity towards the mean, the true measure of the association may have been attenuated.

There is also the potential for residual confounding in the data that may be stronger for these two motor items. For example, children that are very low birth weight are often born with low muscle tone. Low muscle tone in turn affects the child's muscle force production. If children in the treatment group had lower muscle tone than the untreated children, this confounding could have attenuated the true effect of therapy. In this analysis, we analyzed the effect of therapy among children with similar baseline functional ability while controlling for baseline motor score and length of hospital stay after birth among other confounders. However, it is a possibility that our covariates did not capture a proxy for muscle tone, leaving residual confounding in the estimated effect estimates.

To address confounding using measured covariates, we implemented a robust propensity score approach to balance the distribution of known confounders between our treatment groups. To estimate the propensity score, we used a novel method, random forest classification, which has been suggested in simulation studies to achieve better covariate balance and to produce less biased effect estimates when compared with commonly used logistic regression models. However, this approach is rarely used in practice.[58]

In our sample, we first used multiple methods to determine confounders to enter into the propensity score model. To build a parsimonious model in the context of our small sample size, we considered strong confounders in addition to variables that were strongly associated with the outcome. We considered multiple methods to assess confounding: a directed acyclic graph, change in the effect estimate of more than 10% when a confounder was dropped from the model, and the association of confounders and the outcome among children who did not receive the treatment.

We compared covariate balance between treatment groups when the sample was weighted with the propensity score estimated from random forest classification to that of alternative methods used to estimate the propensity score. Our results were similar to that of Lee and colleagues who reported, when compared to standard logistic regression and a single classification tree, random forest and bagging returned the lowest mean absolute standardized differences across these algorithms. [17] In these data, bagging and random forest classification achieved the best overall balance of covariates across treatment groups. The propensity scores estimated from the logistic model showed a marginal imbalance in covariates, where the single classification tree method had the worst performance.

We also considered the results of this estimation method on bias of our estimated effect estimate. The random forest algorithm naturally models higher order effects and non-linear functional forms. Simulation studies suggest that logistic regression main affect models perform adequately in reducing bias when the relationship between the treatment and independent variables is non-linear and non-additive. Yet, the mean absolute bias is up to 30% in the presence of these relationships.[17] The relationship between receiving interventional physical and occupational therapy services and covariates is complex and most likely does not conform to these model assumptions.

In practice, use of these tree based methods, either random forest classification or bagging, are relatively straight forward to implement. The algorithms used to estimate the predicted probability of treatment are executed using the free R statistical platform. Documentation is available for each statistical package (RandomForest,Ipred) to guide the investigator.

In propensity score estimation, both of these tree based methods offer several advantages over estimating with propensity score using traditional logistic regression. These methods are non-parametric where all predictor variables are partitioned in the feature space into rectangular areas where observations have similar responses. This partitioning naturally considers the functional form of the data as well as higher order interactions.[68] When using parametric models, one needs to consider these relationships to prevent model misspecification. For example, if there is strong non additivity between confounders and the predicted probability of treatment, a model with only main effects may misspecify the propensity score model producing a biased effect estimate. For the naïve researcher, both tree based methods require less decision making during model building. One simply enters relevant confounders into the tree based algorithm to estimate the predicted probability of treatment. In simulation studies, when compared with logistic regression, these methods do a better job balancing the distribution of confounders between treatment groups and result in less biased effect estimates.[17]

Second, these methods are useful to estimate predicted probabilities of treatment in the context of small sample sizes where there is the potential for higher order interactive effects. When using traditional parametric models, entering multiple interaction terms in a model with a small sample size may result in cell counts that are too small for parameter estimation.[68] However, tree based methods are advantageous for small sample sizes with large numbers of parameters. At each split, only a select number of random variables are considered. Moreover, each of these variables is considered individually in a sequential order.

This approach was particularly useful in our analysis which included a small sample of approximately 500 VLBW children where the propensity to receive physical or occupational therapy is most likely not linear and may vary within levels of confounding variables: SES for example.

Although these methods offer improvement in propensity score estimation when compared to logistic regression models, the classification made by the algorithm may be difficult to interpret. The algorithm improves the predictive performance by perturbing the data to grow many trees from bootstrap samples.

In addition the number of predictor variables available to split the data at each tree branch is random, therefore decorrelating the tree structure and increasing the variability. However, at the cost of improving prediction accuracy, we are unable to determine the structure of an individual tree as well as the pathway that was taken for classification based on the predictors in the algorithm.[68]

The main limitation of the random forest algorithm, is its potential to overfit the data. In this case, the algorithm mirrors the data too closely, including the true structure of the data as well as the random variation, resulting in poor generalizability to other samples. Breiman addressed this issue reporting that the algorithm did not result in overfitting of the data.[69] However, recent work by Segal in 2004 reported that the algorithm may overfit the data, specifically in the context of deep trees.[103] Whether the algorithm overfits the data and under which parameters is still being researched. The number of variables randomly chosen to split the data in our algorithm was modest (N=4), and our trees were not extensively deep. Based on the findings of the Breiman, we would expect our findings to generalize to other groups of VLBW children with a similar distribution of baseline covariates.

Finally, in our work, we checked the sensitivity of our estimated effect estimates by restricting our population to children in the treated and untreated groups with overlapping propensity score distributions. Moreover, from this overlapping distribution, we then trimmed children who were treated most contrary to prediction based on the 1st and 99th percentile as well as the 2.5th and

97.5th percentile of the distribution. Here, assuming uniformity of effects, there may be an unmeasured confounding factor that influences why an untreated child who should have received the treatment did not and vice versa for the treated group. For example, children who did not receive therapy yet had a propensity to receive treatment that was greater than the 99th percentile were excluded. Children in the treatment group with a propensity for treatment lower than the 1st percentile were also excluded. All of our estimated effect estimates were similar for each of these sensitivity analyses. Thus, we feel our results are robust to unmeasured confounders which may have influence treatment decisions.

In these data, we used robust methods to control for confounding by severity of the child's baseline functional ability. While we limited our cohort to children with similar functional ability at nine months to decrease bias, we greatly reduced our sample size from approximately 2,500 children to 500 children. This decision was a tradeoff between bias and precision. Without restricting our population, the bias due to differential need for therapy may have been too heterogenous to account for with propensity score methods. Yet, restricting our sample size limited our power to detect an effect and decreased the precision of our estimate.

In addition to our small sample size, additional children (approximately 20% of the cohort) were dropped out of the model due to missing data on preschool motor skills. Children who did not have complete information on their ability to perform preschool motor skills were more likely to be from a lower socioeconomic status. Children from a lower socioeconomic group may experience different levels of cognitive stimulation in the home as well as different environmental stressors which may alter their response to therapy treatment. Additional work on disparities in response to treatment may be an interesting avenue for future work in this area.

Overall, methodologically, we used advanced statistical tools and discovered that physical and occupational therapy had a modest positive effect on preschool motor skills. The magnitude of our effect was most likely attenuated and imprecise due to our small sample size, lack of complete data on preschool motor skills, and use of a crude measure of therapy treatment. However, this body of work also has important implications for evidence based clinical practice.

Major developmental milestones among children born with very low birth weight are initially delayed, yet these children show a catch up effect to that of normal birth weight children before or near two years of age.[33] Once these children have reached appropriate age adjusted motor milestones, clinicians may consider discontinuing services. However, these children are at increased risk of motor coordination disorder later in childhood.[10] Providing treatment to these children while ambulation, coordination, and balance skills are emerging may optimize health and function at school age. The implications for improvement in motor performance may have far reaching effects on physical activity, school performance, and self-esteem perhaps into early adulthood.[98,99,100]

In our study, with only a crude measure of services, receipt of physical or occupational therapy services during the toddler years showed a modest, though imprecise effect, on preschool motor coordination. Our misclassification of treatment during this time frame was most likely minimal since treatment plans are typically written for twelve month periods. Yet, our measure did not accurately reflect timing, amount, or duration of therapy which are important parameters when tailoring therapy to improve patient outcomes. One may anticipate that the magnitude of effect may actually be larger than our modest effect, especially with greater intensity or duration of therapy.

This study provides initial support that physical and occupational therapy may be beneficial in improving preschool motor skills among children born VLBW. Yet additional studies should increase the sample size to improve both precision and power of the study as well as collect detailed information on treatment measures. For instance, future work should consider specific parameters of treatment. In the early intervention setting, there is frequent clinical discussion regarding whether to treat children in a clinical setting versus the child's natural environment.[104] In these settings, some children are treated in a group where other therapists treat children one on one. Moreover, there is a move toward an interdisciplinary treatment model. For example, speech therapy and occupational therapy may co treat a child where each therapist focuses on a particular component of a play activity.

In addition to treatment setting, one should also consider the amount of time per week the child receives therapy. It would be helpful for clinicians to understand the dose response relationship between amount of therapy and the child's motor performance. Perhaps there may be a minimal threshold of treatment that the child needs to gain mobility skills and a threshold where additional treatment offers no additional benefit. One must also consider the duration of treatment and how optimal frequency of therapy may vary depending on the length of treatment over time. Understanding these relationships would help to optimize motor function in this very low birth weight population.

In addition to treatment parameters, future studies may also consider the method in which motor performance is evaluated. In our work, we used specific items which assessed motor coordination taken from a standardized assessment. This approach was novel, since the majority of studies in this substantive area measured motor performance using a standardized score. In many of these studies, there was no effect of therapy on infant motor ability.[49,50] However, there is the possibility that therapy is more or less beneficial for targeted skills during different developmental windows. Perhaps there is no difference in motor score by treatment status at follow up, yet children who receive therapy may achieve a given skill earlier in development.

There is the possibility that therapy may differentially impact specific skills, which would be masked using a composite standardized score. Although standardized assessments are useful to compare children to a developmental norm, standard scores may be highly variable in infants and toddlers and the sensitivity of standard scores to detect motor coordination difficulties, especially among children with DCD, has been questioned in the literature.[94-97]

In our work, therapy was beneficial for items involving complex motor coordination (skipping, backwards walking), yet there appeared to be no effect for skills that involved a strength component to the task. Although, standardized assessments provide a score for service eligibility as well as a population percentile reference for children the same age, there may be some benefit to considering either subscale scores or individual tasks. Specifically, future studies may consider attainment of individual skills perhaps using a survival analysis approach.

In this body of work, our sample included children born VLBW with similar functional ability at nine months (successfully cruising), without known congenital anomalies, and who did not appear to have mobility problems with their arms or legs. Due to small cell counts, we were therefore unable to complete an analysis of how our effect may vary within levels of maternal characteristics. However, working in this small sample, we were able to illustrate how we can gain increased precision of our effect estimates when we represent confounders using one scalar, the propensity score, compared to controlling for confounders in a standard logistic regression model. Moreover, this analysis demonstrated the benefits of estimating the propensity score with an ensemble method tree based method such as random forest classification. This method is particularly useful when one has a small sample size with a large number of confounders to be entered into the propensity score model.

In summary, we estimated the effect of physical and occupational therapy services between nine months and two years of age and preschool motor coordination in a sample of very low birth weight children at risk for impaired school age motor coordination. We applied robust methods, inverse probability of treatment weights, among children with similar need for therapy to estimate the average effect of therapy on preschool motor outcomes. We considered several novel methods, random forest classification and bagging, to estimate the predicted probability of treatment. In this sample, very low birth weight children, without congenital anomalies who appeared to be meeting normal milestones at nine months, and who received therapy were more likely to be successful with complex motor coordination compared to children who did not receive the treatment. Although our magnitude of the effect was modest and imprecise, most likely due to our small sample size and crude estimate of treatment, these findings provide initial support that providing therapy services may attenuate preschool motor coordination impairment among children born VLBW.

Although modest, these findings provide initial support for evidence based treatment of children born VLBW. Even though additional work is needed to more clearly define initiation, duration, and timing of therapy, these findings support the general delivery of early intervention physical and occupational therapy services to facilitate preschool motor performance in these children.

Although VLBW children may catch up to their normal birth weight peers during their toddler years, these children are at risk for poor coordination later in childhood. Healthcare reimbursement organizations and policy makers should consider these findings when making decisions regarding reimbursement and program guidelines for early intervention services for children born VLBW.

Clinically, these results support the treatment of VLBW children during the toddler years to facilitate motor control as new complex motor skills are beginning to emerge. Additional research is needed to fine tune the parameters of treatment in this population as well as the effect of therapy among children born VLBW with more severe neurological impairments.

Table 1. Skilled Physical/Occupational/Physiotherapy Intervention and Motor Outcomes in LBW Preterm Children

Author	Year	Study Design	Population	Sample Size	Exposure	Frequency	Timing of Exposure
Goodman et al	1985	RCT	<1700 grams, GA<34 weeks	N=107	Outpatient NDT	1xmox45minx12mo	PT in the first year of life
Piper et al	1986	RCT	<1500 grams	N=115	Physiotherapy	Every 2 wks for 9 mo	From 40 wks postmentrual age
Barrera et al	1986		<1500 grams, 1500-2000 grams	N=20(<1500 grams) N=39 (1500 to 2000grams)	Group 1: Developmental therapy (training in speech and OT) ; Group 2: educational training of parent	1-2 hrs wkly X3mo, 2Xmo 6 mo, quarterly until 1 year	From 4 months CA
Rothberg et al	1991	RCT	<1700 grams, <34 weeks, NICU stay	N=25 (normal) N=24 (at risk)	Treated monthly at hospital with home program	45 minutes session in the hospital	From 3 months- 12 months
Girolami and Campbell	1994	RCT	<35 weeks and <1800 grams with 3 abnormal reflexes on NBAS at 34- 35 weeks postconception	N=19	Developmental PT	14-28 treatment session 2Xday 15 mins	34-35 weeks post conception over 7 to 17 days
Salokorpi	1998	RCT	<1000 grams	N=104	OT NDT and SI	60 minutes per week	6mo (adjusted age) until 12 months (adjusted age)
Lekskulchai and Cole	2001	RCT	GA<37 weeks	N=34 control, N=38 intervention	Developmental PT intervention	Monthly	Term age until four months corrected age
Yigit	2002	RCT	<34 weeks and <2,000 grams considered "low risk"	N=160	Physical Therapy	1xmo 9 months and every other month until 18-24 mo	Birth until age 2
Salokorpi et al	2002	RCT	<1000 grams	N=126	OT NDT and SI	60 minutes per week	6mo (adjusted age) until 12 months (adjusted age)
Cameron	2005	RCT	<32 weeks, <1500 grams	N=72	Neonatal developmental program delivered by pediatric physiotherapists		Early PT intervention
Koldewijn	2010	RCT	<32 weeks and/or <1500 g	N=176	Infant Behavioral Assessment and Intervention Program	6-8 sessions delivered by a pediatric PT	Post discharge intervention until 6 months

Table 1 Cont. Skilled Physical/Occupational/Physiotherapy Intervention and Motor Outcomes in LBW

Author	Age at Assessment	Outcome Measures	Main Effects	Subgroup Analyses
Goodman et al	12 mo	Griffith GCI and locomotor subscales	No significant difference in DQ score or locomotor subscale score between intervention and control for either "at risk" or "normal" low birth weight infants	No
Piper et al	6 months, 12 months	Wolanski Gross Motor Evaluation, Milani Comparetti Motor Development Screening Test, Griffith Mental Development Scale	No significant differences on any dependent measure of motor performance between intervention and controls at 6 months or 12 months	No
Barrera et al	16 months	Bayley Mental and Motor Score	Lower Bayley scored or VLBW compared to LBW. No sig differences in motor scores between interv and control for either VLBW or LBW groups	Infants <1500 grams showed greater gains with developmental intervention in the Bayley Mental Score than infants 1500<2000
Rothberg et al	12 months adjusted, two assessments before age 3, and assessment age 6	Griffith Mental Development Scale, Griffith 2 Scale ages 2-8	No difference between physiotherapy and control groups for "normal" or "at risk" children at either 1 year or 6 years	No
Girolami and Campbell	NICU Discharge	Neonatal Behavioral Assessment Scale, Supplmental Motor Test	SMT exp group has statistically sign higher scores on functional postural scores. Tx group demonstrated more midlien behaviors, anti gravity movement, and head rotation. No differences noted on the Neonatal Assessment Scale.	No
Salokorpi	3, 6, 9 , 12, 18 ,24 months		No difference in neurodevelopment with exception of social development at 12 months	No
Lekskulchai and Cole	4 months corrected age	Test of Infant Motor Performance	Intervention group scored signficantly higher on the TIMP than the control group at four months corrected age	No
Yigit	throughout	Reflexes and motor milestones		No
Salokorpi et al	age 4	Miller Assessment for Preschoolers	No significant difference in any MAP sub scores between groups	No difference between groups within strata of neurological status: minor vs not minor impairment at either age 2 or age 4. NO STATISTICALLY SIGN interactions with GA, dev risk score, gender
Cameron	4 months corrected age	AIMS Assessment at 4 months	No significant difference between Median AIMS percentile rank between the two groups at 4 months	No
Koldewijn	24 months corrected age	BSID-II, CBCL	Children in the intervention has an increase of 6.4 points ± 2.4 p=0.006 on the motor scale. This estimate was adjusted for perinatal and background variables	Subgroup analyses showed improved motor and mental outcomes in intervention infants with bronchopulomonary dysplasia and combined biological and social risk factors

Table 2. Other Interventions and Developmental Outcomes

Low Birth Weight Population and Developmental Outcomes

Author	Year	Study Design	Population	Sample Size	Exposure	Frequency	Timing of Expsure
Resnick	1987	RCT	LBW 500 to 1800 grams	N=255	Infant Developmental Program (visual, vestibular stim, PROM) by nursing, OT, psychologist through age 2	NICU and twice a month after discharge	NICU until age 2
Infant Health and Development Program	1990	RCT	Low Birth weigt (<2500 grams) premature (<37 week) infants	N=985	Home visits, child attendance at a child development center, parent group meetings	Learning activities at child development center 5 days a week	Hosppital discharge through 36 months
Gianni	2006	RCT	<1250 grams	N=36	Mother Child Intervention Program: mother support and promotion of perceptual and social cognitive skills	Twice a month	3 to 12 months
Ho et al	2010	RCT	25- 35 wks and <1500 grams	N=24	Massage therapy in the NICU 15 mins	5xwk for 4 wks	NICU
Hamilton et al	1999	Quasi	At risk for Developmental Delay N=11 preterm,SGA	N=43	Parent Led Motor Skills		Post Test 8 wk follow up

Normal Birth Weight Population and Motor Outcomes

Author	Year	Study Design	Population	Sample Size	Exposure	Frequency	Timing of Expsure
Riethmuller	2009	Systematic Review	Children < 5 yo	10 published studies	Motor Development Interventions: delivered by researchers, teachers, parents: non skilled	5-30 hours of instruction	Average 11 weeks

Table 2 Cont. Other Interventions and Developmental Outcomes

Low Birth Weight Population and Developmental Outcomes

Author	Age at Assessment	Outcome Measures	Main Effects	Subgroup Analyses
Resnick	12 and 24 months	Bayley Scales	Experimental group had significantly higher scores mental and physical indices. Lower orevalence of developmental delay in experimental group	No
Infant Health and Development Program	36 months	Cognitive competence (Stanford Binet), Behavioral competence (CBC), health status (morbidity, functional status, matern perception of child health).	Intervention group had significantly higehr mean IQ scores. No difference in mean score or Functional Status II Scale.	Among healvier babies, Stanford Binet scores were significantly higher for intervention kids. Small statistically significant behavioral advantage for intervention among less educated mothers
Gianni	36 months CA	Griffiths Mental Development Scale	No difference in subscale scores at 12 and 24 months. At 36 months children in the intervention had higher mean scores in eye hand coordination,personal social, and practial reasoning skills	No
Ho et al	Pre Assessment 34 weeks and post at 38 weeks;	Test of Infant Motor Performance	No difference in TIMP score gain between groups	Looked at differences among kids by GA, BW, TIMP score at 34 weeks. Among babies with an initial low TIMP score (<35 points) the intervention group showed sign higher TIMP score p=0.043
Hamilton et al	Post Test 8 wk follow up	TGMD:object control skills	Statistically sign change in object control total score in the exp group…not change observed in the control group	No

Normal Birth Weight Population and Motor Outcomes

Author	Age at Assessment	Outcome Measures	Main Effects	Subgroup Analyses
Riethmuller		Motor Skills	7 studies found statistically significant improvements in motor skills. There was variability in the skills that they assesses: locomotor, object control	No

Table 3. ECLS-B samples sizes for children's' receipt of services*

	Normal	Low	Very Low	Total
9-Month Services				
Physical therapy	50	50	250	300
Vision services	30	0	150	150
Hearing services	0	0	50	50
Social work services	50	50	150	200
Psychological services	0	0	50	50
Home Service	50	50	250	300
Parent support/training	50	50	100	150
Any services	150	100	350	450
2-Year Services				
Speech therapy	100	50	200	250
Occupational therapy	50	50	150	200
Physical therapy	50	50	200	250
Vision services	0	0	100	100
Hearing services	0	0	50	50
Social work services	50	0	50	50
Psychological services	0	0	0	0
Home Service	100	50	200	250
Parent support/training	50	0	50	50
Special needs class with other children	0	0	50	50
Any services	150	100	250	350
Preschool Services				
Speech therapy	250	100	200	300
Occupational therapy	100	50	150	200
Physical therapy	50	50	150	200
Vision services	50	50	100	150
Hearing services	50	50	50	100
Social work services	50	50	50	100
Psychological services	50	0	50	50
Home Service	100	50	100	150
Parent support/training	100	50	50	100
Special needs class with other children	150	50	100	150
Private tutoring	50	0	50	50
Instruction in Braille	0	0	0	0
Instruction in Sign-language	0	0	0	0
Any services	300	100	200	300
Kindergarten Services				
Speech therapy	150	50	100	150
Occupational therapy	100	50	100	150
Physical therapy	50	50	100	150
Vision services	50	0	50	50
Hearing services	0	0	0	0

Social work services	50	0	50	50
Psychological services	100	0	50	50
Home Service	50	0	50	50
Parent support/training	50	0	0	0
Special needs class with other children	100	50	50	100
Private tutoring	50	0	50	50
Instruction in Braille	0	0	0	0
Instruction in Sign-language	0	0	0	0
Any services	250	100	150	250

*Estimates rounded to nearest 50

Table 4. Comparison of baseline characteristics by early childhood physical (PT) or occupational therapy (OT) receipt among low birth weight children

N≈3,000		Received early childhood PT/OT (N≈300)	Did not receive early childhood PT/OT (N≈2400)	P value
Gender				
	Male	59.70%	46.65%	<0.0001
	Missing		0.21%	
Race/Ethnicity				
	White Non- Hispanic	51.12%	46.61%	0.2817
	African American	20.15%	23.55%	
	Hispanic	18.66%	18.58%	
	Asian	2.24%	4.27%	
	Other	7.84%	6.82%	
	Missing		0.16%	
Parental Education				
	Less than High School	11.94%	14.84%	0.7989
	High School or equivalent	26.49%	25.73%	
	Technical Training/Some college	30.22%	28.81%	
	Bachelor's Degree	17.54%	17.02%	
	Graduate Professional Training	13.43%	13.52%	
	Missing	0.37%	0.08%	
SES level (mean;std)		-0.101 (0.818)	-0.120 (0.846)	0.7248
Urbanicity				
	Urban, inside UA	71.27%	71.19%	0.4047
	Urban, inside UC	14.18%	11.92%	
	Rural	14.55%	16.89%	
Region				0.6133

Northwest	16.79%	16.73%	
Midwest	26.87%	23.51%	
South	36.57%	39.87%	
West	19.78%	19.89%	
Number of Siblings			0.9703
0	33.96%	32.10%	
1	33.96%	33.99%	
2	19.03%	20.43%	
3	8.21%	8.59%	
4 or more	4.85%	4.89%	
Birth Weight	1137.36 (494.89)	1785.25 (573.80)	<0.0001
Gestational age; wks (mean,std)	29.05 (4.30)	33.67 (4.45)	<0.0001
Missing	3.73%	1.6%	
5 minute AGPGAR score; (mean, std)	7.42 (1.47)	8.38 (1.18)	<0.0001
Missing	12.31%	16.35%	
Length of hospital stay; days (mean,std)	75.09 (53.11)	23.68 (32.16)	<0.0001
Missing	0.37%	0.16%	
Length of NICU admission; days (mean, std)	5.68 (19.83)	2.76 (11.97)	0.0191
Missing	-	0.25%	
Health Insurance (Not Mutually Exclusive)			
Private	52.24%	50.84%	0.6445
SCHIP	9.09%	9.95%	0.6460
Medicaid	60.45%	46.44%	<0.0001
Military Insurance	1.49%	2.47%	0.3216
Indian Health Services	-	0.21%	0.0777
Other	6.72%	2.55%	0.0001
Missing	0.37%	0.21%	

Health condition at birth* (Not Mutually Exclusive)			
Visual Deficit	25.37%	2.92%	<0.0001
Heart Defect	14.93%	5.51%	<0.0001
Difficulty hearing	10.07%	0.78%	<0.0001
Failure to Thrive	11.19%	1.52%	<0.0001
Difficulty with mobility	31.34%	2.47%	<0.0001
Cleft palate	1.49%	0.41%	0.0191
Other Special Need	32.84%	3.82%	<0.0001
Missing	0.37%	0.16%	
Frequency of treatment for injury ; birth and 9 months			
Never	98.88%	95.51%	0.5240
Once	3.37%	4.03%	
Twice	0.75%	0.29%	
Three or More	-	0.16%	
Overall caregiver health status			
Excellent	21.27%	29.92%	0.0068
Very Good	35.07%	32.31%	
Good	29.85%	27.21%	
Fair	11.19%	9.49%	
Poor	2.61%	0.95%	
Missing	-	0.13%	
Age adjusted Bayley Motor T score (mean,std)	35.48 (14.09)	46.56 (10.23)	<0.0001
Missing	16.0%	6.58%	
HOME score (mean,std)	5.30 (1.21)	5.47 (1.12)	0.0395
Missing	5.60%	-	

KIDI score (mean,std)	6.84 (2.04)	6.65 (2.14)	0.1389
CESD score (mean,std)	6.70 (6.37)	5.66 (5.83)	0.0151
Missing	7.84%	11.55%	
Partner in the home			
Yes	73.13%	74.72%	0.2018
Caregiver social support#			
Yes	99.25%	98.77%	0.6193
Missing	-	0.16%	
Age when child sat without support (months)			
Not sitting independently	45.52%	6.70%	<0.0001
2 to 4	0.37%	3.21%	
5 to 7	12.31%	55.49%	
8 to 14	41.42%	34.48%	
Missing	0.37%	0.12%	
Age when child crawled on hands and knees (months)			<0.0001
Not crawling	64.55%	21.70%	
3 to 6	1.12%	15.74%	
7 to 9	15.67%	49.98%	
10 to 14	18.28%	12.49%	
Missing	0.37%	0.08%	
Age when pulled to standing (months)			<0.0001
Not pulling to stand	73.51%	25.98%	
4 to 7	1.87%	15.91%	
8 to 10	13.43%	48.86%	
11 to 14	10.82%	9.54%	
Missing	0.37%	0.12%	

Age walked while holding onto something (months)			<0.0001
Not cruising	77.61%	36.87%	
5 to 7	0.37%	7.44%	
8 to 10	10.45%	42.99%	
11 to 14	11.19%	12.58%	
Missing	0.37%	0.12%	
Maternal work schedule			0.0197
Mother does not work	61.57%	51.34%	
Regular daytime shift	27.24%	35.14%	
Regular evening shift	5.60%	5.67%	
Regular night shift	2.61%	1.81%	
Rotating shift	4 (1.49%)	71 (2.92%)	
Split shift	-	0.78%	
Other	1.12%	2.10%	
Missing	0.37%	0.25%	

*Sixteen kids with congenital defects affecting motor ability were excluded from the analysis. Defects included Downs Syndrome, Turners Syndrome, and Spina Bifida. #Caregivers were asked about the people they turn to for social support and advice with childcare. Among the 3,000 low birth weight children, 8.19% of the sample were missing data on receipt of therapy between 9 months and age 2. We used a chi squared test (for categorical data) or independent t tests (assuming unequal variance between groups) to test for significant differences in covariates by therapy status.

Table 5. Candidate variables for predictors of therapy services between 9 months and 2 years of age

Health Insurance	Categorical	General Knowledge of child development	Continuous
9 month child’s health condition	Categorical	9 month Parenting measure (KIDI)	Continuous
Birth weight	continuous	9 month maternal CESD score	Continuous
9 month Bayley Motor T score	continuous		
Race/ethnicity	categorical	9 month Injury in the last three months	dichotomous
Parental education	categorical	9 month caregiver health	categorical
Income	continuous	9 month HOME score	continuous
Gestational Age at Birth	continuous		
Number of Siblings	continuous	9 month Zip code	continuous
5 minute APGAR scores	continuous	SES scale	continuous
Length of hospital stay	continuous	Paternal Education	
Length of NICU stay	continuous	9 month hours per week in childcare	continuous
Mother’s occupation	Categorical	Urbanicity	dichotomous
9 month parent report of age of motor milestone attainment	Sitting, crawling, pulling to stand, walking	Geographic region	Nominal

Table 6. Variable importance of preschool impairment in motor coordination among low birth weight children [@*$]

Variable	MDC*	Variable	MDC*	Variable	MDC*
Length Hospital Stay(days)	0.65	Congenital Deficit	0.17	Child's Weight;2 years	0.03
Unable to stand	0.63	Sat without support 7 to 14 month	0.17	Sat without support 2 to 4 months	0.03
Birth weight (grams)	0.55	APGAR score	0.16	Breastfed > 19 mo	0.03
Gestational Age(weeks)	0.54	CESD score	0.14	Excellent caregiver health	0.03
Unable to cruise	0.48	Fair caregiver health	0.14	Difficulty with UE/LW mobility	0.02
Unable to crawl	0.47	Crawled 10 to 14 mo	0.14	Cruised 10 to 14 months	0.00
Cruised 7 to 10 months	0.45	Child injured once birth to 9 mont	0.14	Child injured >3 times birth to 9 months	0.00
Bayley Motor T score	0.44	Hours of weekly TV; age 2	0.14	Military health insurance	-0.01
Weekly childcare hours; 2 yrs	0.41	Good caregiver health	0.13	Breastfed ≤11mo	-0.01
SES level	0.40	Failure to thrive	0.13	Neighborhood safety	-0.01
Crawled 3 to 7 months	0.37	Days in NICU	0.12	Child injured twice; birth to 9 months	-0.04
Weekly childcare hours at 9months	0.37	Bachelor's Degree	0.12	Poor caregiver health	-0.04
Private Health Insurance	0.35	Maternal partner in home	0.11	Indian Health Insurance	-0.06
Sat without support 4 to 7 months	0.32	Ever breastfed	0.10	Maternal Support	-0.07
Maternal work schedule	0.29	***Cruised 10 to 14 months***	***0.10***	Other Health Insurance	-0.08
Crawled 7 to 10 months	0.27	Other special healthcare need	0.09		
Unable to sit	0.26	Less than High School Education	0.08		
Pulled to stand 8 to 11 months	0.26	High School Education	0.07		
Pulled to stand 4 to 8 months	0.26	HOME score	0.07		
Graduate Training	0.25	Mother Ever Major Depression	0.07		
Race	0.24	Visual Deficit	0.07		
PT/OT receipt	0.23	SCHIP insurance	0.07		
Medicaid insurance	0.22	Did not breastfeed	0.07		
KIDI score	0.22	Very good caregiver health	0.06		
HOME score; 2 years	0.21	Number of ear infections	0.06		
Number of siblings	0.20	Technical/some college	0.06		
Urbanicity	0.19	Hearing deficit	0.06		
Mother works	0.19	Breastfed 11 to 19 months	0.04		
Region	0.18	Pulled to stand 11 to 14 months	0.03		

* The mean decrease in accuracy (MDC) is the difference in classification accuracy using the out of bag data when the variable is included and the classification accuracy when the values of the variable in the out of bag variable are permuted randomly. A higher mean decrease accuracy score indicates a variable of greater importance in prediction of preschool impairment in motor coordination. [&] Variable importance measures estimated using RandomForest package with 500 trees and 9 variables randomly chosen at each node. [@]Variables with a mean decrease in accuracy above the absolute value of the lowest mean decrease in accuracy estimate is suggested as a cut point for importance in predicting class membership of the outcome. Therefore variables with a mean decrease in accuracy above 0.08 are suggested important predictors of impaired preschool motor ability.

Table 7. Marginal relationship between continuous/ordinal covariates and the exposure[*, ^, #]

Variable	Shape of the relationship
Number of hours per week in childcare	Generally linear
APGAR scores	U shaped
KIDI score	U shaped
SES level	U shaped
Gestational Age	U Shaped
Birth weight	Relatively linear
Number of days the child stayed in the hospital after birth	Strong Linearity
9 month Age adjusted BSF-R Motor T score	U shaped
Maternal Depression CESD score	Linear

[*]Partial dependence plots represent the marginal effect of one variable on receipt of early childhood PT/OT therapy services. The plot provides the average trend of the variables after averaging out the effects of the other predictor variables in the model. ^Plots generated using the RandomForest package with 500 trees and 6 variables chosen randomly at each split. Variables included parent education, child injury, caregiver health, maternal support, child congenital deficit, visual deficit, difficulty with upper or lower extremity mobility, region, urbanicity, maternal education, father education, health insurance, race, hearing impairment, failure to thrive, other special healthcare need, maternal work schedule, maternal partner in home, age sitting, age crawling, age standing, age walking, birth weight, gestational age, SES level, KIDI score, CESD score, HOME score, 9 month BSF-R Motor R score, APGAR, length of hospital stay after birth, number of hours per week in childcare

Table 8. Association between covariates and therapy services and ability to pass skipping item in the unexposed

Variable	Early Childhood Motor Performance			Preschool Skipping		
	OR	95% CI	CLR	OR	95% CI	CLR
Baseline(9mo)						
Health Insurance: Dichotomous variables						
Medicaid						
Yes	1.34	0.84,2.13	2.54	1.01	0.60,1.70	2.83
Private Health Insurance						
Yes	1.50	0.94,2.40	2.55	1.03	0.61,1.74	2.85
Children's State Health Plan						
Yes	0.67	0.30,1.51	5.03	1.85	0.88,3.92	4.45
Child's Health Status						
Visual Deficit	4.54	2.44,8.45	3.47	0.77	0.22,2.70	12.27
Hearing Deficit	0.24	0.09,0.69	7.62	No events here	Drop	
Other special healthcare need	0.21	0.12,0.39	3.21	5.44	0.72,40.91	56.82
Failure to thrive	0.34	0.11,1.11	11.14	No events here	Drop	
Birth weight (grams): Scaled at 200 grams	0.73	0.62,0.86	1.39	0.93	0.79,1.09	1.38
<1000 grams	Ref	-	-	Ref	-	-
1000-1500 grams	0.54	0.34,0.87	2.56	0.93	0.53,1.62	3.06
>1500 grams	0.20	0.06,0.68	11.33	0.58	0.22,1.51	6.86
Bayley Motor T Scores scaled 10 unit inc	0.55	0.43,0.69	1.61	1.01	0.76,1.34	1.76
Race/ethnicity						

White	ref	-	-	ref	-	-
Black	0.53	0.29,0.97	3.34	3.18	1.75,5.78	3.30
Other	0.74	0.42,1.29	3.07	1.04	0.49,2.20	4.49
Parental Education						
<High School	ref	-	-	ref	-	-
High School	0.97	0.42,2.30	5.48	1.80	0.77,4.23	5.49
Some College	1.23	0.56,2.71	4.84	1.22	1.86,2.76	1.48
College/Graduate Training	1.94	0.92,4.08	4.43	1.07	0.47,2.45	5.21
Gestational Age at Birth(wks)	0.86	0.79,0.93		1.01	0.93,1.09	1.17
Number of Siblings	0.98	0.81,1.18	1.46	0.78	0.57,0.97	1.65
5 minute APGAR score	0.85	0.73,0.99	1.36	0.89	0.74,1.08	1.45
Length of Hospital stay(days)						
30 to 45 days	ref	-	-	ref	-	-
45 to 60 days	2.72	0.92,8.08	8.78	0.74	0.33,1.64	4.97
60 to 90 days	4.85	1.88,12.49	6.64	0.83	0.45,1.53	3.40
>90 days	9.25	3.33,25.73	7.73	1.02	0.37,2.82	7.62
Parent report of age of motor milestone attainment (categorical)						
Age at which the child sat indep	1.47	1.27,1.71	1.35	0.97	0.83,1.15	1.35

Age at which the child crawled	1.62	1.32,1.97	1.49	1.00	0.82,1.20	1.47
Age at which the child walked	1.35	1.07,1.71	1.60	0.83	0.66,1.05	1.59
Parental knowledge of child development (KIDI score)						
<5	ref	-	-	ref	-	-
5 to <10	0.97	0.49,1.94	3.96	0.61	0.30,1.25	4.17
>= 10	0.56	0.18,1.72	9.56	0.96	0.37,2.49	6.73
Maternal depression (CESD score) 5 unit increase	1.08	0.87,1.32	1.51	1.08	0.85,1.37	1.57
General health of the caregiver						
Excellent	0.46	0.25,0.84	3.36	1.11	0.57,2.14	3.41
Fair	0.62	0.26,1.47	5.65	1.87	0.76,4.63	6.09
Good	0.57	0.32,1.01	3.16	0.86	0.43,1.86	3.78
Poor	0.54	0	63.66	1.25	0.13,11.78	90.62
Very Good	ref	-	-	ref	-	-
Caregiver received advice with childcare						
Cognitive stimulation in the home (HOME score) 9 months						

<2		Too small				
2<=4	1.01	0.52,1.87	3.60	1.01	0.51,2.04	4.00
5-6	ref	-	-	ref	-	-
>6	1.89	0.92,3.89	4.23	0.63	0.21,1.87	8.90
Region						
Northwest	1.00	0.48,2.09	4.35	1.56	0.64,3.81	6.24
Midwest	0.90	0.45,1.79	3.97	1.02	0.44,2.35	5.34
South	0.85	0.46,1.57	3.39	1.10	0.52,2.35	4.52
West	ref	-	-	ref	-	-
Urbanicity						
Urban inside UA	0.93	0.50,1.75	3.50	1.35	0.63,2.92	4.63
Urban inside UC	0.59	0.23,1.56	6.78	1.42	0.53,3.81	7.19
Rural	ref	-	-	ref	-	--
Hours per wk in childcare 9mo	1.00	0.99,1.01	1.02	0.99	0.98,1.01	1.03
<10	ref	-	-	ref	-	-
10 to <=20	0.58	0.20,1.68	8.40	0.90	0.33,2.49	7.54
20 to <=40	0.82	0.45,1.52	3.38	1.43	0.75,2.72	3.63
>40	0.75	0.36,1.53	4.25	0.56	0.23,1.46	6.35
Work schedule: collapse this variables						
Not working	1.67	0.57,4.88	8.55	3.29	0.75,14.43	19.24
Regular days	1.25	0.41,3.80	9.27	2.63	0.58,11.82	20.38
Other	ref	-	-	ref	-	-
Partner in the home	No variability					
SES						
<-0.70	ref	-	-	ref	-	-
-0.70 to 0.56	1.67	0.87,3.22	3.68	1.07	0.57,1.99	3.33
>0.56	2.37	1.18,4.74	4.02	0.76	0.35,1.63	4.66

Table 9. Therapy exposure and preschool skipping ability: change in estimate

Backwards Elimination	Beta coefficient	Upper 95 CI%	Lower 95% CI	OR	Upper 95% CI	Lower 95% CI	CLR	ln(OR full/OR reduced)
Full Model: therapy,sitting,standing, cruising, race,visual deficit, hospital, education, caregiver health, gestwk,medicaid, private,schip, ses, homescore, workschedule,region,kidiscore	0.6563	-1.2527	2.5654	*1.92764683*	0.285732277	13.00585967	45.517643	
Motor score	0.6415	-1.2569	2.5399	1.899327735	0.284534718	12.67840307	44.558369	0.0148
Birth weight	*0.5223*	*-1.3564*	*2.401*	*1.685900766*	*0.257586421*	*11.03420507*	*42.836905*	*0.134*
Age of sitting	0.6531	-1.2917	2.5978	1.921488219	0.27480322	13.43415037	48.886437	0.0032
Age of crawling	0.7328	-0.8167	2.2822	2.080898975	0.44188748	9.798212783	22.173547	-0.0765
Age of cruising	*0.3228*	*-1.0457*	*1.6914*	*1.380989125*	*0.351445721*	*5.427073292*	*15.442138*	*0.3335*
SES	0.6136	-1.2779	2.505	1.847068892	0.278621792	12.24355897	43.943293	0.0427
Time in hospital after birth	*0.2393*	*-1.6096*	*2.0881*	*1.270359587*	*0.199967585*	*8.06956841*	*40.354382*	*0.417*
Race	*0.3802*	*-1.3753*	*2.1357*	*1.462577076*	*0.252763755*	*8.462968512*	*33.481733*	*0.2761*
KIDI score	0.6515	-1.2513	2.5543	1.918416296	0.286132583	12.8622929	44.952213	0.0048
HOME	*0.5093*	*-1.3686*	*2.3872*	*1.664125899*	*0.254462958*	*10.8829789*	*42.768421*	*0.147*
Education	*0.344*	*-1.4274*	*2.1153*	*1.410578636*	*0.239931935*	*8.292073015*	*34.560106*	*0.3123*
Medicaid	*0.3176*	*-1.5397*	*2.175*	*1.373826621*	*0.214445425*	*8.802185122*	*41.046271*	*0.3387*
SCHIP	0.6393	-1.2656	2.5442	1.895153807	0.282070003	12.73303758	45.14141	0.017
Private HI	0.5498	-1.2653	2.365	1.732906402	0.282154637	10.64403882	37.724132	0.1065
Visual Deficit	0.5408	-1.2732	2.3549	1.717380217	0.279934397	10.5370751	37.64123	0.1155
Caregiver health	0.6986	-1.2274	2.6246	2.010935426	0.293053527	13.79905346	47.087143	-0.0423
Gestational Age	0.6561	-1.2534	2.5656	1.927261339	0.285532335	13.00846111	45.558627	0.0002
Region	0.7672	-1.1266	2.6609	2.153727367	0.324133439	14.30916155	44.145897	-0.1109
Work Schedule	0.8011	-1.0144	2.6166	2.22799037	0.362619937	13.68910143	37.750548	-0.1448
APGAR	*0.3299*	*-1.419*	*2.0787*	*1.390829039*	*0.241955852*	*7.994069866*	*33.039374*	*0.3264*

Table 10. Candidate variables for the propensity score model

Variable	Minimally Sufficient Conditioning Set: DAG	Covariates associated with skipping	Change in estimate >10%
Age of cruising	*X*	*X*	*X (strong)*
Age crawling			
Age of independent sitting			
Gestational Age	*X*	*X*	
Medicaid	x	X	x
Days in the hospital after birth	X	X	X
Birth weight	X	X	X
SES	X	X	
Race	X	X	X
Education	X	X	X
Vision	x	x	x
Other Special healthcare need	x	x	
Motor Score	X	X	
KIDI score		x	
Work Schedule		x	x
APGAR		X	x
Private HI		x(marginal)	
Caregiver health		x	
Number of siblings		x	
HOME score	x		x
Hours in childcare	x	x	

Table 11. Sensitivity of the effect estimates in the propensity score model^

N≈700	Skipping				Hopping				Balance				Walking backwards			
	OR	Lower 95% CI	Upper 95% CI	CLR	OR	Lower 95% CI	Upper 95% CI	CLR	OR	Lower 95% CI	Upper 95% CI	CLR	OR	Lower 95% CI	Upper 95% CI	CLR
Propensity Score 1*																
Stabilized	*3.16*	*1.02*	*9.83*	*9.67*	*2.01*	*0.75*	*5.42*	*7.27*	*1.56*	*0.69*	*3.50*	*5.06*	*0.86*	*0.32*	*2.31*	*7.30*
Propensity Score 2%																
Stabilized	*2.58*	*0.97*	*6.90*	*7.13*	*1.21*	*0.55*	*2.69*	*4.92*	*1.59*	*0.76*	*3.29*	*4.31*	1.15	0.50	2.66	5.35

^ Propensity estimated using random forest classification with 1,000 trees and 7 variables randomly chosen at each node; models adjusted for birth weight, length of hospital stay, age at which the child cruised *Variable strongly associated with outcome and DAG: GA, SCHIP Health Insurance, Visual Deficit, Other Special Healthcare need, birth weight, race, education number of siblings, APGAR score, child's length of hospital stay after birth, age of cruising, HOME score, region, maternal work schedule, hours in childcare, SES % Reduction of propensity score model to strong confounders to improve precision: GA, Birth weight, race, education, length of the child's hospital stay after birth, age of cruising, SES, BSF-R Motor T score

Table 12. Association between receipt of physical and occupational therapy in early childhood and preschool motor performance among very low birth weight children ^ *

	Crude Estimate			Standard Model Adjusted Model		
Preschool Motor Skills[&]	*N*	*OR*	*95 % CI*	*N*	*OR*	*95 % CI*
Skipping eight consecutive steps[&]	500	0.75	0.25,2.28	300	1.67	0.46,6.03
Hopping five times independently[&]	500	0.50	0.21,1.17	300	0.80	0.30,2.11
Maintaining single leg stance for ten seconds independently[&]	500	0.77	0.34,1.72	350	0.82	0.33,2.04
Walking Backwards six steps on a line[&]	500	0.86	0.34,2.22	350	1.04	0.37,2.88
		Beta(SE)	P Value		Beta (SE)	P Value
Jumping Distance (inches) [&]	500	-1.39 (1.72)	0.42	350	0.33(1.86)	0.86
Change in Jumping Distance Preschool-Kindergarten (inches)[#]	500	-0.18 (2.39)	0.94	400	-1.02 (2.70)	0.70

^Early Childhood Longitudinal Study, Birth Cohort 2001-2006. All counts have been rounded to the nearest 50 for data security. [$] We defined children who received physical and occupational therapy between nine months and age two as treated. * The model was missing the outcome for ≈20% of the sample. [&] The model was adjusted for length of hospital stay(continuous). cruising (continuous), birth weight(continuous), gestational age (continuous), race, education, SES (restricted quadratic spline), 9 month BSF-R Motor T score (continuous) [#] Model adjusted for preschool jumping distance and age of cruising (continuous)

Table 13. Descriptive statistics of VLBW children by receipt of therapy services between nine months and age two^

N≈500	Received early childhood PT/OT (6.5%)	Did not receive early childhood PT/OT (83.3%)
	%	%
Gender		
Male	61.3	45.3
Missing	-	<.8%
Race/Ethnicity		
White Non- Hispanic	58.1	38.5
African American	19.4	33.2
Hispanic	16.1	20.5
Other	6.5	7.9
Parental Education		
Less than High School	12.9	18.2
High School or equivalent	19.4	29.4
Technical Training/Some college	35.5	26.1
Bachelor's/Graduate Degree	32.3	26.1
Missing	-	<0.8%
SES level (mean;std)	-0.14(0.71)	-0.27 (0.80)
Urban city		
Urban, Inside Urban Area	64.5	68.9
Urban, Inside Urban Cluster	12.9	12.9
Rural	22.6	18.2
Region		
Northwest	19.4	14.2
Midwest	19.4	20.8
South	41.9	48.1
West	19.4	17.0
Number of Siblings		
0	54.8	43.3

1	25.8	27.3
2	16.1	18.5
3	3.2	6.8
4 or more	-	4.1
Birth Weight (Mean;std)	1030.23 (212.63)	1147.77(250.31)
Gestational age; wks (mean,std)	27.84(3.17)	30.13(3.80)
Missing	-	1.8%
Fetal Growth Ratio[&]	0.94(0.28)	0.78(0.30)
Missing	-	3.8%
Age adjusted *BSF-R T score* [#]	45.15 (10.85)	49.81 (8.44)
Missing	<9.7%	3.8%
5 minute AGPGAR score; (mean, std)	7.15 (1.89)	7.89 (1.26)
Missing	16.1%	14.9%
Length of hospital stay; days (mean,std)	67.94(22.33)	48.46(30.40)
Health Insurance (Not mutually exclusive)		
Private	54.8	40.3
SCHIP	6.5	12.2
Medicaid	64.5	54.9
Military Insurance	-	3.5
Indian Health Services	-	0.8
Other	<9.7	3.3

Health condition at birth* (Not mutually exclusive)		
Visual Deficit	19.4	6.1
Heart Defect	12.9	8.4
Difficulty hearing	6.5	1.3
Failure to Thrive	6.5	0.8
Cleft palate	-	0.8
Other Special Need	19.4	4.3
Frequency of treatment for injury ; birth and 9 months		
Never	83.9	96.2
Once	12.9	3.5
Twice	3.2	<0.8
Overall caregiver health status		
Excellent	29.0	26.6
Very Good	35.5	31.7
Good	19.4	29.6
Fair	16.1	10.4
Poor	-	1.8
HOME score (mean,std)	5.70(1.09)	5.44(1.15)
Missing	<9.7%	2.3%
KIDI score (mean,std)	6.81(1.80)	6.46(1.97)
CESD score (mean,std)	5.0(4.98)	5.66(5.73)
Missing	9.7%	10.9%

Partner in the home		
Yes	74.2	69.9
No Partner	3.2	12.2
Caregiver social support		
Yes	96.8	99.8
Age when child sat without support (months)	8.45 (1.57)	7.77(1.73)
Missing	<9.7%	1.0%
Age when child crawled on hands and knees (months)	9.81(1.63)	8.56(1.74)
Missing	16.1%	6.8%
Age walked while holding onto something (months)	11.00(1.53)	(10.00)1.59
Maternal work schedule		
Mother does not work	54.8	52.2
Regular daytime shift	35.5	35.7
Regular evening shift	9.7	5.6
Regular night shift	-	3.0
Rotating shift	-	1.8
Split shift	-	1.3
Other	-	0.5
Pre K Mean Jumping Distance inches	21.12(8.58)	22.52(8.41)
Missing	16.1%	16.6%
Pre K Ability to Maintain SLS 10 seconds	35.5%	40.3%
Missing	12.9%	15.7%
Child did not respond	3.2%	2.0%

Pre K Ability to Hop 5 times	35.5%	50.6%
Missing	12.9%	15.7%
Child did not respond	12.9%	6.1%
Pre K Ability to Walk Backwards on line for six steps	19.4%	21.0%
Missing	16.1%	15.4%
Child did not respond	-	3.0%
Pre K Ability to Skip 8 alternating steps	12.9%	16.5%
Missing	16.1%	16.0%
Child did not respond	9.7%	9.1%

[*]Therapy status missing for 10.13% of the sample. All counts have been rounded to the nearest 50 according to the data use agreement. Counts with less than three cases are expressed as approximate percentages according to data use agreement. [#]The *BSF-R T score* has a mean of 50 and standard deviation of 10. [&]Fetal growth ratio=birth weight/median birth weight for gestational age. ^ Early Childhood Longitudinal Study, Birth Cohort 2001-2006

Table 14. Standardized difference among confounders

Confounder	Unweighted			Random Forest		
	Early Childhood Therapy Mean(SE)	No Early Childhood Therapy Mean(SE)	Standardized Differences^	Early Childhood Therapy Mean (SE)	No Early Childhood Therapy Mean(SE)	Standardized Differences^
Birth weight (grams)	1030.23 (212.63)	1147.77 (250.31)	0.51	1095.65 (746.04)	1139.93 (259.76)	0.08
BSF-R T score	45.16 (10.85)	49.81 (8.44)	0.48	49.00 (32.91)	49.44 (8.88)	0.02
Days in Hospital after birth	67.94 (22.33)	48.46 (30.40)	0.73	58.30 (87.00)	49.27 (31.78)	0.14
SES	-0.14 (0.71)	-0.27 (0.80)	0.17	-0.27 (2.66)	-0.25 (0.84)	0.01
Age of independent sitting (months)	8.45 (1.57)	7.77 (1.73)	0.41	7.83 (6.53)	7.84 (1.82)	0.00
Age of crawling (months)	9.81 (1.63)	8.56 (1.74)	0.74	9.12 (7.62)	8.62 (1.79)	0.09
Age of walking with support (months)	11.00 (1.53)	10.00 (1.59)	0.64	10.26 (4.96)	10.07 (1.66)	0.05
Gestational Age (weeks)	27.84 (3.17)	30.13 (3.80)	0.65	28.78 (11.16)	30.04 (3.92)	0.15
Mean			*0.54*			*0.07*

^Standardized differences represent the differences between means by therapy status in units of standard deviations. The estimates are calculated as $d=(xbar_{therapy}-xbar_{no\ therapy})/sqrt(S^2_{therapy} + S^2_{no\ therapy})/2))$[75]

Table 15. Association between receipt of physical and occupational therapy in early childhood and preschool motor performance among VLBW children: Average Treatment Effect[^ $ *]

	Crude Estimate (N≈500)			Adjusted Estimate (N≈300)		
Preschool Motor Skills[&]	N	OR	95 % CI	N	OR	95 % CI
Skipping eight consecutive steps[&]	300	0.85	0.28,2.63	300	2.39	0.75,7.51
Hopping five times independently[&]	300	0.56	0.23,1.32	300	0.90	0.33,2.45
Maintaining single leg stance for ten seconds independently[&]	350	0.74	0.32,1.71	350	1.07	0.45,3.13
Walking Backwards six steps on a line[&]	350	1.01	0.39,2.63	350	1.52	0.51,4.54
	N	Beta	95% CI	N	Beta	95% CI
Jumping Distance (inches) [&]	350	-1.12	-4.65,2.40	350	1.79	-2.21,5.79
Change in Jumping Distance Preschool-Kindergarten (inches)[#]	300	0.10	-4.86,5.07	300	-0.76	-4.45,3.69

^ Early Childhood Longitudinal Study, Birth Cohort 2001-2006. All counts have been rounded to the nearest 50 for data security. These data were weighted using (1/propensity score) for children who received therapy and (1/(1-propensity score)) for those children who did not receive treatment. The propensity scores were stabilized by multiplying the treatment weights by the marginal prevalence of the treatment that they actually received. [$] We estimated the propensity score using random forest classification. The out of bag error rate for the algorithm was 15.65% across 1,000 trees where the algorithm chose 4 random variables at each split of the node. [&] Weighted models were adjusted for age at which the child walked (continuous), birth weight (continuous), number of days the child was in the hospital after birth (continuous term). [*] We defined children who received physical and occupational therapy between nine months and age two as treated. 7% of children were missing covariate data to estimate the propensity score and are not included in the final mode. Outcome data were missing for approximately 20% of the sample, 6% of children were missing data on receipt of therapy. [#] Model adjusted for preschool jumping distance and age of walking with support (continuous)

Table 16. Sensitivity Analysis: Overlapping propensity scores and propensity scores trimmed contrary to prediction

	*Overlapping Propensity Score***		*1st and 99th % trimmed*		*2.5 and 97.5 % trimmed*&&	
	OR	*95 % CI*	*OR*	*95 % CI*	*OR*	*95 % CI*
Preschool Motor Skills&						
Skipping eight consecutive steps& (N≈300)	*2.27*	0.71,7.29	*2.27*	0.71,7.29	2.22	0.69,7.11
Hopping five times independently& (N≈300)	0.86	0.32,2.36	0.86	0.32,2.36	0.83	0.32,2.46
Maintaining single leg stance for ten seconds independently& (N≈350)	1.13	0.43,3.00	1.13	0.43,3.00	1.15	0.43,3.06
Walking Backwards six steps on a line& (N≈350)	1.43	0.47,4.31	1.43	0.47,4.31	1.57	0.52,4.77
	Beta (SE)	P Value	Beta (SE)	P Value	Beta (SE)	P Value
Jumping Distance (inches) & (N≈350)	1.99 (2.11)	0.35	1.99 (2.12)	0.35	2.09(2.11)	0.32

^These data were weighted using (1/propensity score) for children who received therapy and (1/(1-propensity score)) for those children who did not receive treatment. The propensity scores were stabilized by multiplying the treatment weights by the marginal prevalence of the treatment that they actually received. &We estimated the propensity score using random forest classification. The out of bag error rate for the algorithm was 15.65% across 1,000 trees where the algorithm chose 4 random variables at each split of the node $ Weighted models were adjusted for age at which the child walked (continuous), birth weight (continuous), number of days the child was in the hospital after birth (continuous term). #We defined children who received physical and occupational therapy between nine months and age two as treated. **25% of sample trimmed && Trimmed an additional 2% of the overlapping sample

Table 17. Out of bag error rates* for prediction of receipt of early childhood therapy

		Out of Bag Error	
Number of Trees	250	750	1000
Random Forest	%	%	%
Number of Randomly Chosen Variables per split			
2	15.42	15.42	15.42
4	15.42	15.42	15.65
7	15.87	16.10	16.10
Bagging	17.01	16.55	15.65

*The out of bag data is put down each bootstrap classification tree and the results are aggregate to determine the out of bag error rate over the forest of trees.

Table 18. Distribution of propensity score and weights for the average treatment effect by method used to estimate the propensity score*

	Random Forest Classification		Logistic Regression		Classification Tree		Bagging	
	Early Childhood Therapy	No Early Childhood Therapy	Early Childhood Therapy	No Early Childhood Therapy	Early Childhood Therapy	No Early Childhood Therapy	Early Childhood Therapy	No Early Childhood Therapy
Propensity Score&								
Minimum	0.0133333	0	0.0232836	0.0015716	0.0606061	0	0.0081744	0
Maximum	0.4150943	0.5350877	0.4278358	0.5646912	0.6363636	0.6363636	0.5482094	0.5856354
Mean	0.1609705	0.0607887	0.1590819	0.0653820	0.3873993	0.0453312	0.2024614	0.0670286
Average Treatment Effect Weights								
Minimum	0.1575566	0.9345992	0.1528644	0.9360703	0.1027728	0.9345992	0.1192990	0.9345992
Maximum	4.9050634	2.0102699	2.8088781	2.1469798	1.0791139	2.5701477	8.0007032	2.2554993
Mean	0.8859185	1.0025157	0.7852070	1.0086340	0.2706242	1.0044533	0.9911689	1.0147520

*Average treatment effect weight is estimated as (1/propensity score) for those children who received early childhood therapy. For the those children who did not receive therapy, the weight is (1/(1-propensity score)). These weights are stabilized so the sum of the weights reflects the size of the original population. We multiplied the weight by the probability of receiving the treatment that the child actually received. & The propensity score includes the following covariates: 9 month BSF-R motor T score, socioeconomic status, length of child's hospital stay after birth, gestational age, birth weight, parental education, race, age at which the child walked with assistance.

Table 19. Standardized differences among confounders by propensity score method

	Unweighted			Random Forest Classification		
Confounder	Early Childhood Therapy	No Early Childhood Therapy	Standardized Differences^	Early Childhood Therapy	No Early Childhood Therapy	Standardized Differences^
	Mean(SE)	Mean(SE)		Mean (SE)	Mean(SE)	
Birth weight (grams)	1030.23 (212.63)	1147.77 (250.31)	0.506118959	1095.65 (746.04)	1139.93 (259.76)	0.079270498
BSF-R Motor T Score	45.16(10.85)	49.81(8.44)	0.478983337	49.00 (32.91)	49.44(8.88)	0.018086429
Days in Hospital after birth	67.94(22.33)	48.46(30.40)	0.730335059	58.30(87.00)	49.27(31.78)	0.137873049
SES	-0.14(0.71)	-0.27(0.80)	0.172342152	-0.27 (2.66)	-0.25(0.84)	0.012114035
Age of independent sitting (months)	8.45(1.57)	7.77(1.73)	0.409912102	7.83 (6.53)	7.84(1.82)	0.001916724
Age of crawling (months)	9.81(1.63)	8.56(1.74)	0.743953372	9.12 (7.62)	8.62(1.79)	0.09016314
Age of cruising (months)	11.00 (1.53)	10.00(1.59)	0.642874112	10.26 (4.96)	10.07(1.66)	0.050855437
Gestational Age (weeks)	27.84 (3.17)	30.13 (3.80)	0.653973416	28.78 (11.16)	30.04 (3.92)	0.151260955
Mean			0.542311564			0.067692533

^Standardized differences represent the differences between means by therapy status in units of standard deviations. The estimates are calculated as $d=(\bar{x}_{therapy}-\bar{x}_{no\ therapy})/\mathrm{sqrt}(S^2_{therapy} + S^2_{no\ therapy})/2))$[75]

Table 19. Cont. Standardized differences among confounders by propensity score

	Logistic Regression			Classification Tree		
Confounder	Early Childhood Therapy	No Early Childhood Therapy	Standardized Differences^	Early Childhood Therapy	No Early Childhood Therapy	Standardized Differences^
	Mean(SE)	Mean(SE)		Mean (SE)	Mean(SE)	
Birth weight (grams)	1047.03(673.98)	1138.16(262.88)	0.178147366	1054.48(418.10)	1142.43(257.16)	0.253107791
BSF-R Motor T Score	48.34(34.29)	49.36(8.85)	0.040491907	47.79 (17.99)	49.33(8.95)	0.108566495
Days in Hospital after birth	60.97(66.65)	49.60(32.32)	0.21712781	59.41 (47.86)	49.61(31.50)	0.241978953
SES	-0.26(2.46)	-0.25(0.85)	0.004364264	-0.11(1.29)	-0.23(0.85)	0.115194857
Age of independent sitting (months)	8.11(5.76)	7.85(1.83)	0.059293677	7.56(3.52)	7.83(1.80)	0.09764106
Age of crawling (months)	9.45(5.95)	8.63(1.81)	0.186545326	9.22(4.01)	8.61(1.78)	0.198614043
Age of walking with assistance (months)	10.44(4.42)	10.08(1.67)	0.109149723	10.38(2.89)	10.07(1.65)	0.131168368
Gestation Age weeks)	29.28 (13.57)	29.99 (3.93)	0.070610258	28.45 (6.11)	30.04 (3.90)	0.310029113
Mean			0.108216291			0.182037585

^Standardized differences represent the differences between the means by therapy status in units of standard deviations. The estimates are calculated as $d=(\bar{x}_{therapy}-\bar{x}_{no\ therapy})/\mathrm{sqrt}(S^2_{therapy} + S^2_{no\ therapy})/2))$[75]

Table 19. cont. Standardized differences among confounders by propensity score method

	Bagging		
	Early Childhood Therapy	No Early Childhood Therapy	Standardized Differences^
	Mean (SE)	Mean(SE)	
Birth weight (grams)	1146.48(673.17)	1138.75(262.42)	0.015130296
BSF-R Motor T Score	48.53(31.48)	49.35(9.00)	0.035697109
Days in Hospital after birth	51.26(85.37)	49.58(31.97)	0.026119109
SES	-0.25(2.92)	-0.24(0.85)	0.004642274
Age of independent sitting (months)	7.55(7.04)	7.85(1.84)	0.059484057
Age of crawling (months)	8.63(8.22)	8.64(1.82)	0.000862572
Age of cruising (months)	10.10(5.19)	10.09(1.68)	0.004984611
Gestational Age (weeks)	29.43 (11.17)	30.02 (3.95)	0.07058523
Mean			0.027188157

^Standardized differences represent the differences between the means by therapy status in units of standard deviations. The estimates are calculated as $d=(xbar_{therapy}-xbar_{no\ therapy})/sqrt(S^2_{therapy} + S^2_{no\ therapy})/2))$[75]

Table 20. Average treatment effect of interventional physical or occupational therapy services and preschool motor skills: using three methods to estimate the propensity for treatment[^ $ *]

Preschool Motor Outcomes	Crude Model				Random Forest Classification				Logistic Regression			Bagging		
Preschool Motor Skills[&]	N	*OR*	*95 % CI*	*CLR*	N	*OR*	*95 % CI*	*CLR*	*OR*	*95% CI*	*CLR*	*OR*	*95 % CI*	*CLR*
Skipping eight consecutive steps[&]	500	*0.75*	*0.25,2.28*	*9.26*	*300*	*2.39*	*0.75,7.51*	9.87	2.35	0.71,7.71	10.78	3.44*	1.04,11.37	9.78
Hopping five times independently[&]	500	*0.50*	*0.21,1.17*	*5.48*	*300*	*0.90*	*0.33,2.45*	7.48	0.96	0.34,2.69	7.95	1.16	0.41,3.26	7.97
Maintaining single leg stance for ten second independently[&]	500	*0.77*	*0.34,1.72*	*5.03*	*350*	*1.07*	*0.45,3.13*	6.92	0.93	0.34,2.53	7.35	1.70	0.63,4.58	7.24
Walking Backwards six steps on a line[&]	500	*0.86*	*0.33,2.22*	*6.63*	*350*	*1.52*	*0.51,4.54*	8.96	1.50	0.49,4.59	9.36	1.88	0.88,7.37	8.39

^ Early Childhood Longitudinal Study, Birth Cohort 2001-2006. Counts rounded to the nearest 50 according to data use agreement. These data were weighted using (1/propensity score) for children who received therapy and (1/(1-propensity score)) for those children who did not receive treatment. The propensity scores were stabilized by multiplying the treatment weights by the marginal prevalence of the treatment that they actually received. [&] Models were adjusted for age at which the child walked (continuous), birth weight (continuou s), number of days the child was in the hospital after birth (continuous term). *p<0.05 [$]We defined children who received physical and occupational therapy between nine months and age two as treated.*7% percent of children were missing at least one covariate used to generate the propensity score and were excluded. Outcome data were missing for approximately 20% of the sample, 6% of children were missing data on receipt of therapy

Figure 6. Flow Diagram of sample

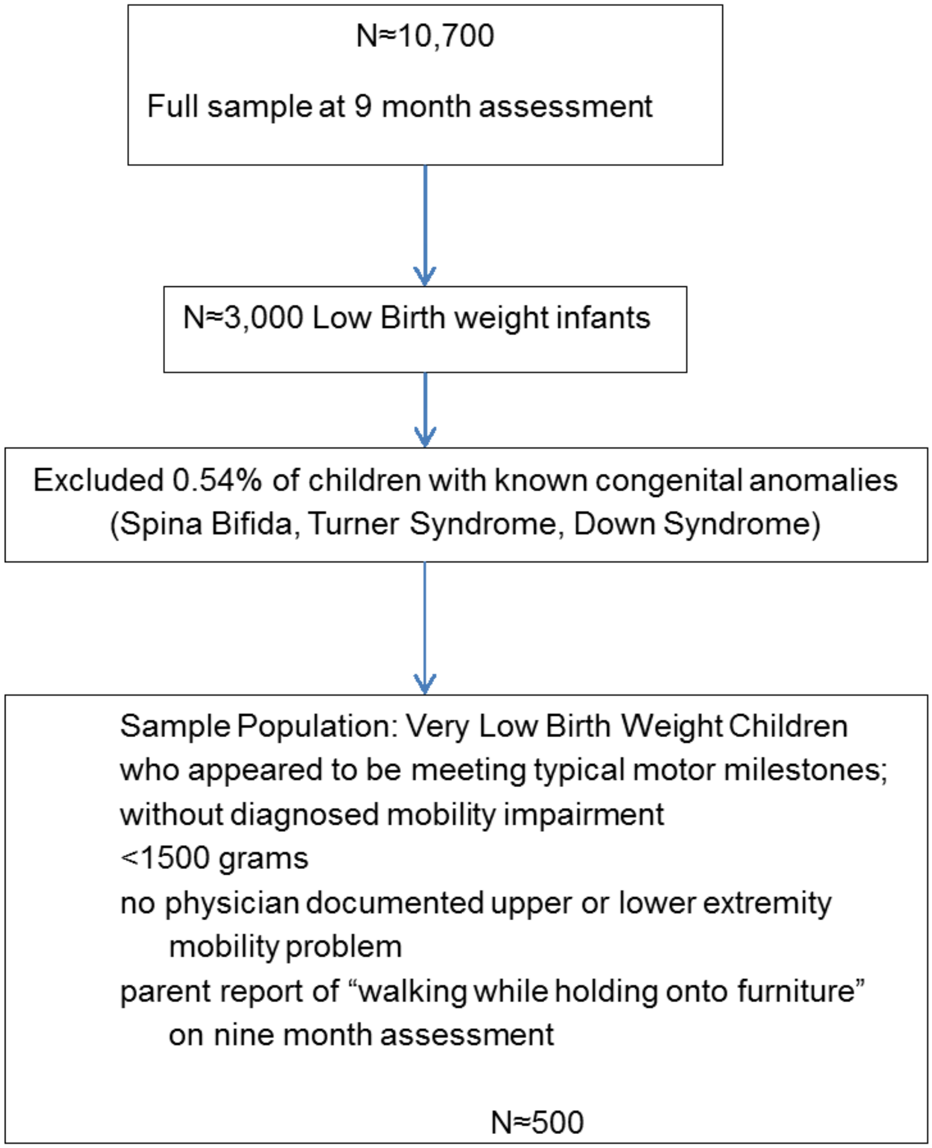
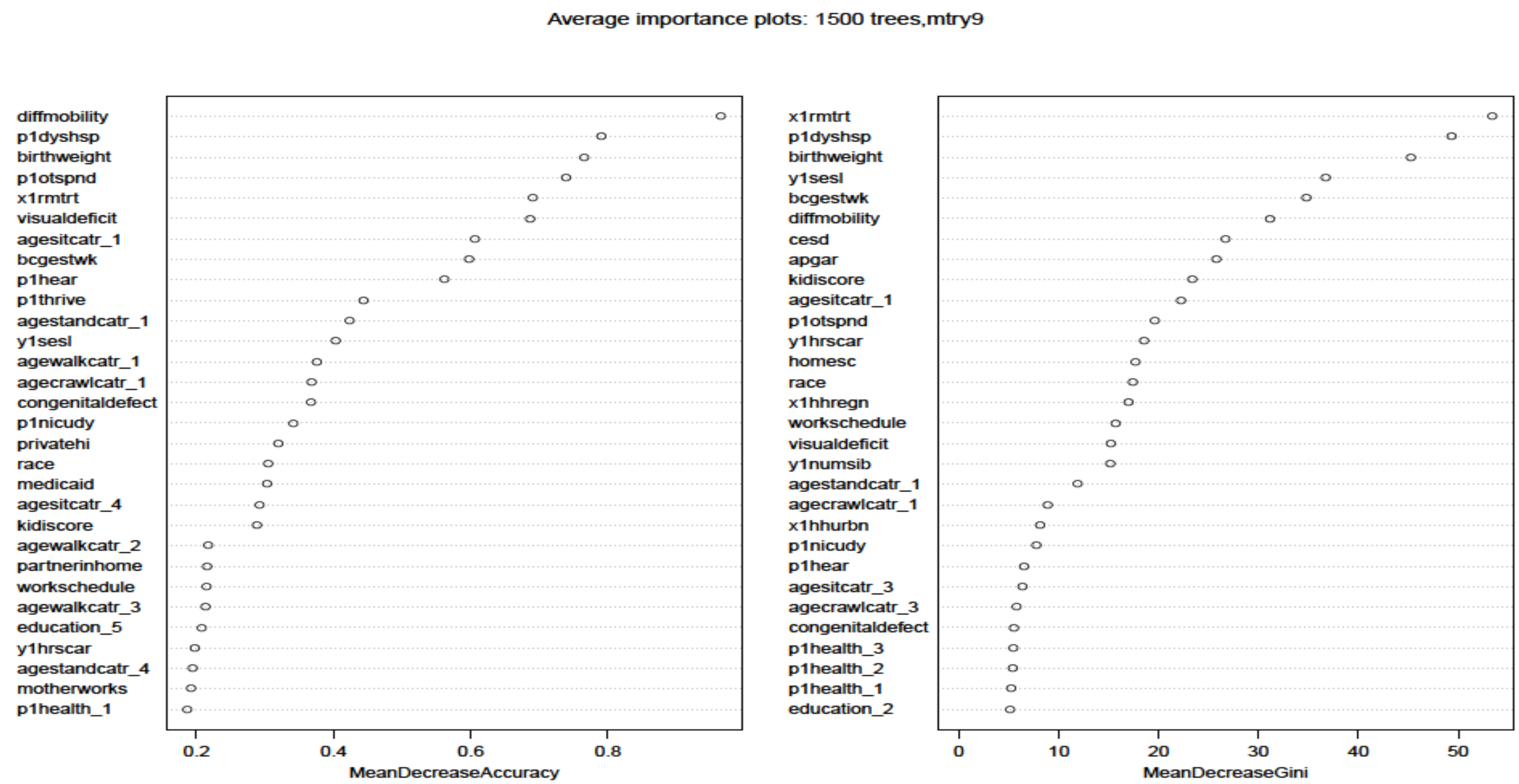
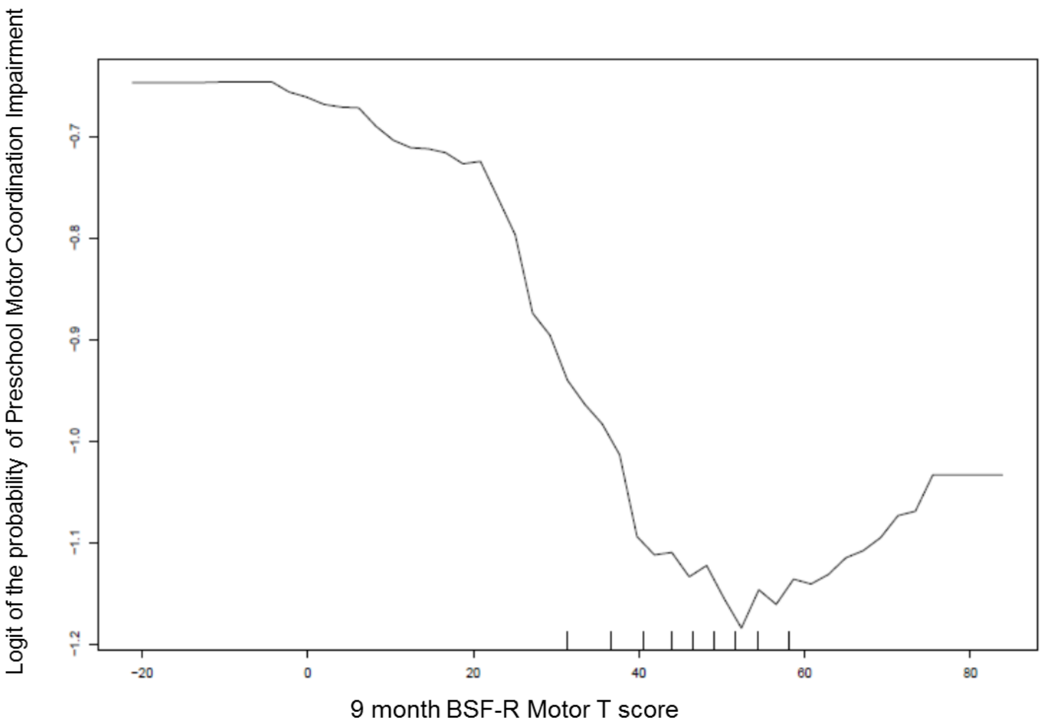


Figure 7. Variable importance in predicting receipt of physical or occupational therapy between 9 months and age 2 among low birth weight infants



* The mean decrease in accuracy is the difference in classification accuracy using the out of bag data when the variable is included and the classification accuracy when the values of the variable in the out of bag variable are permuted randomly. A higher mean decrease accuracy score indicates a variable of greater importance in prediction of receipt of early childhood therapy services.

Figure 8. Partial Dependence Plot between 9 month BSF-R Motor T score and Impaired Preschool Motor Coordination among Low Birth Weight Children



Variable in Random Forest classification model:9 month Bayley Motor T Score, Weekly childcare hours, number of siblings, length of hospital stay, length of NICU stay, hearing deficit, failure to thrive, other special healthcare need, gestational age, ever breastfed, number of ear infections, weekly hours of television watching, neighborhood safety, maternal depression, birth weight, APGAR score, HOME score, KIDI score, region, urbanicty, health insurance, maternal work schedule, PT/OT receipt, maternal support, congenital deficit, visual deficit, difficulty with upper or lower extremity mobility, childhood injury, caregiver health, age at sitting, crawling, standing, cruising, parental education

Figure 9. Partial Dependence Plot of birth weight and impaired preschool motor coordination among low birth weight children

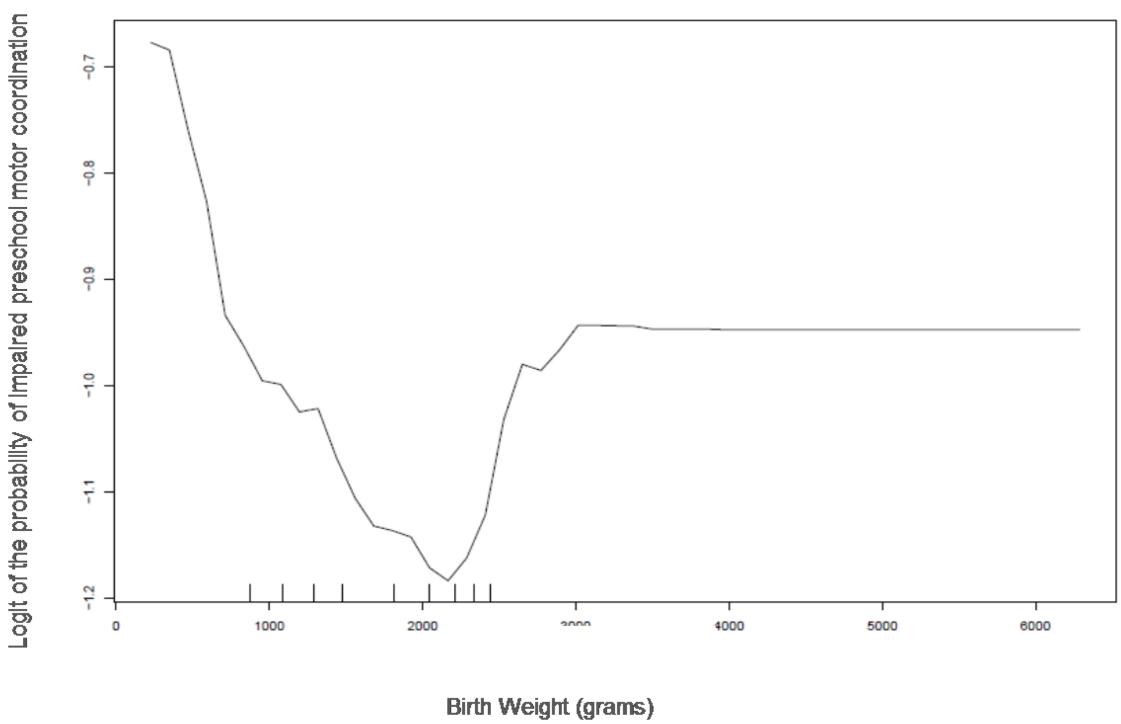


Figure 10. Partial dependence plot of SES and impaired preschool motor coordination among low birth weight children

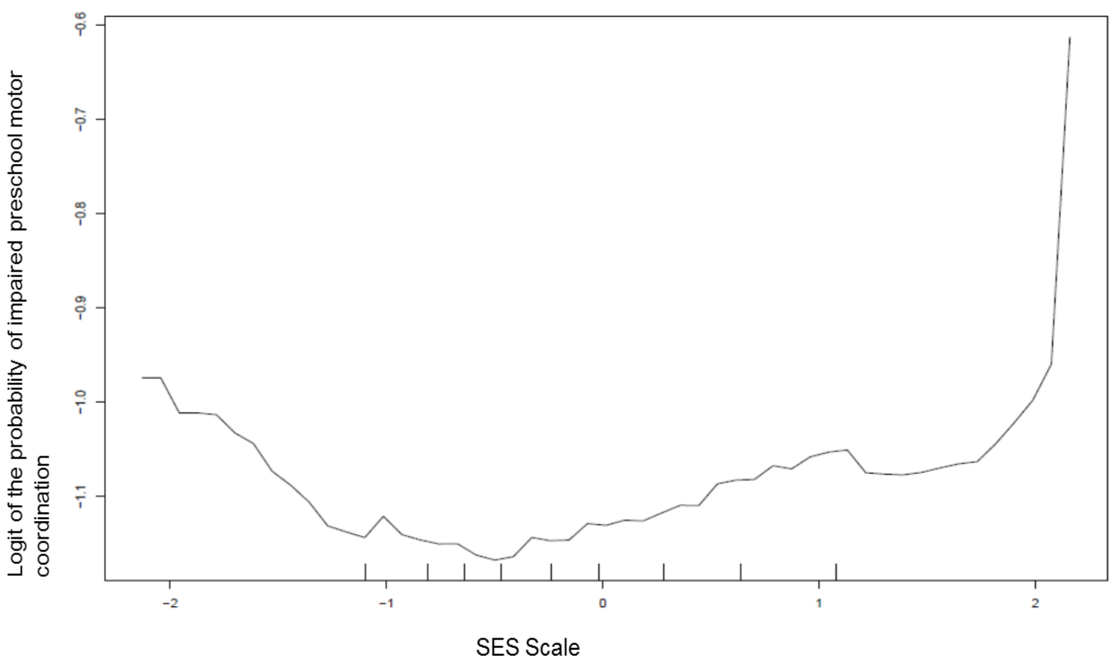


Figure 11. Partial dependence plot between gestational age and impaired preschool motor coordination among low birth weight children

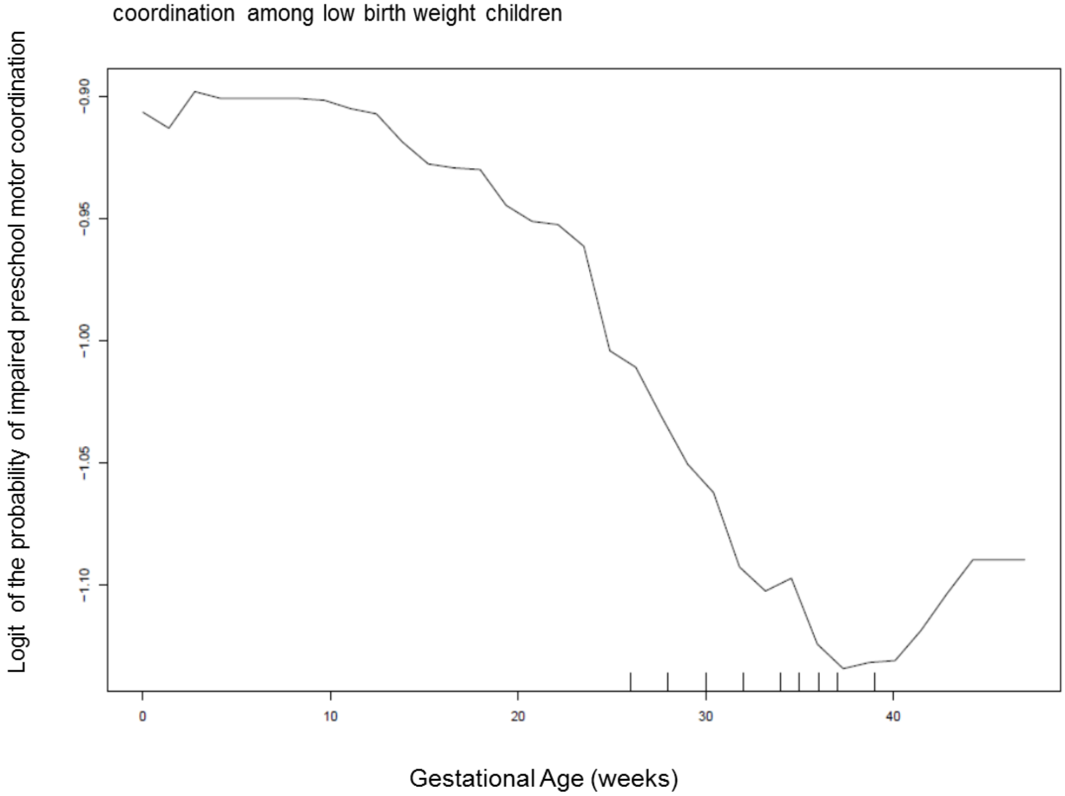


Figure 12. Partial dependence plot between length of hospital stay and impaired preschool motor coordination among low birth weight children

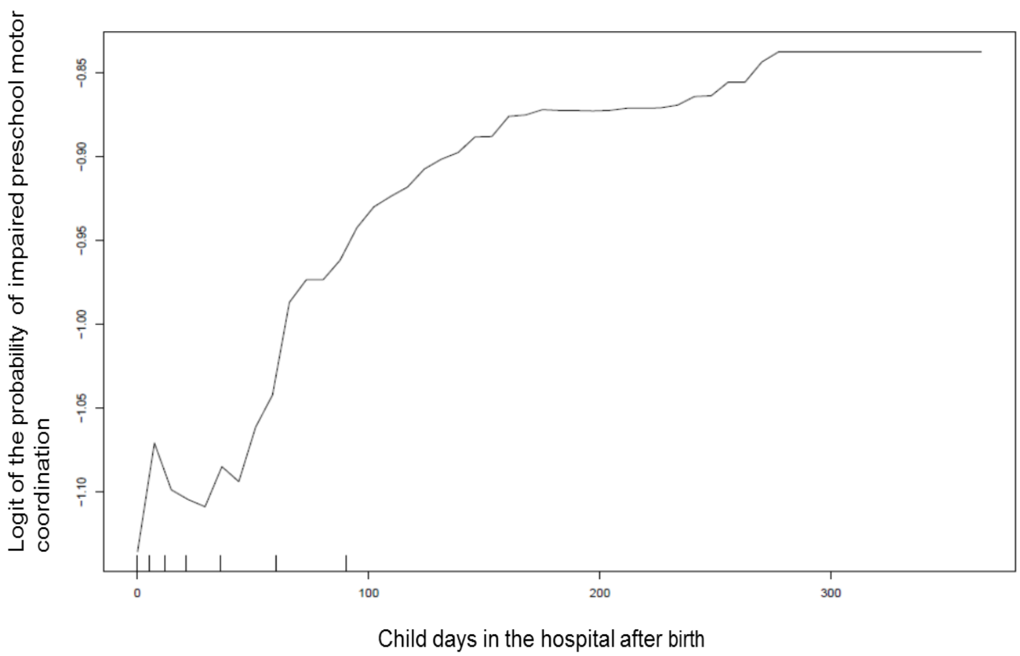


Figure 13. Partial Dependence plot for gestational age and receiving interventional physical or occupational therapy among low birth weight children

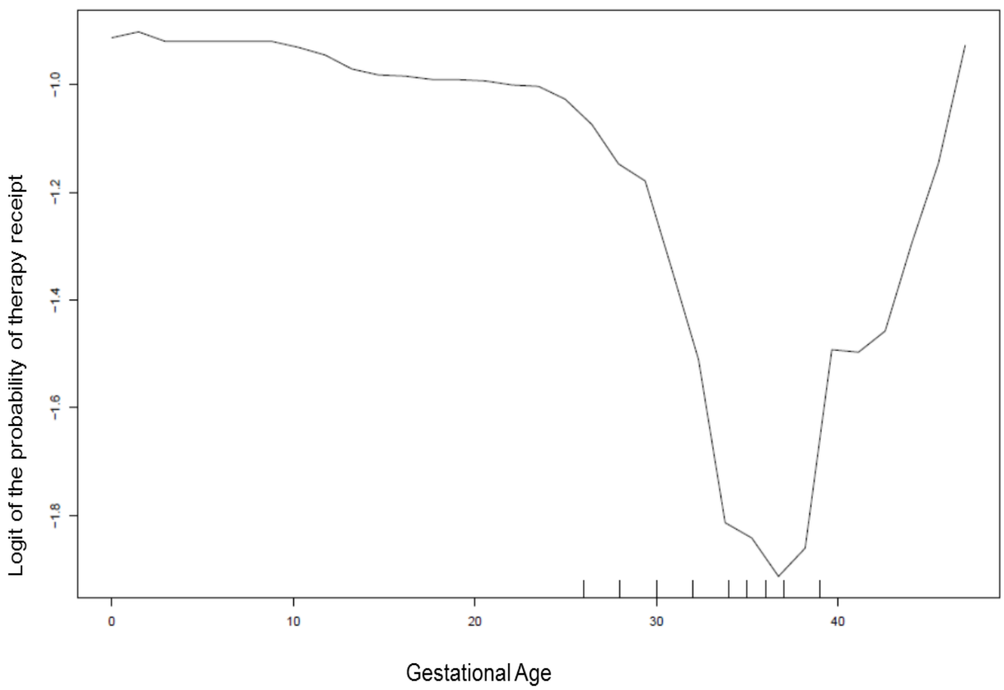


Figure 14. Partial Dependence plot between birth weight and receiving interventional physical or occupational therapy among low birth weight children

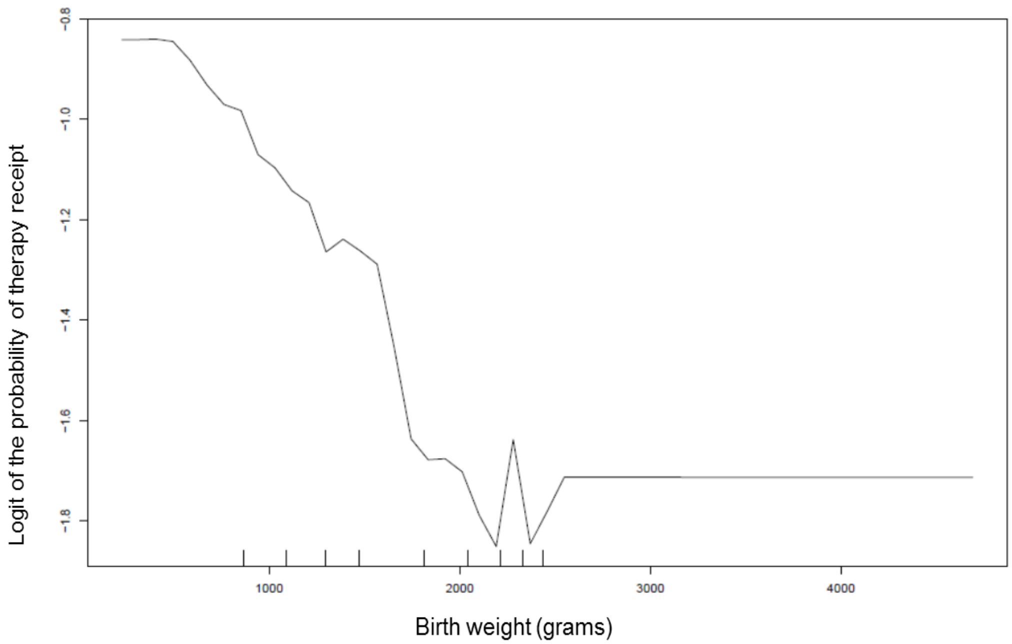


Figure 15. Partial Dependence plot between BSF-R motor T score and receiving interventional physical or occupational therapy among low birth weight children

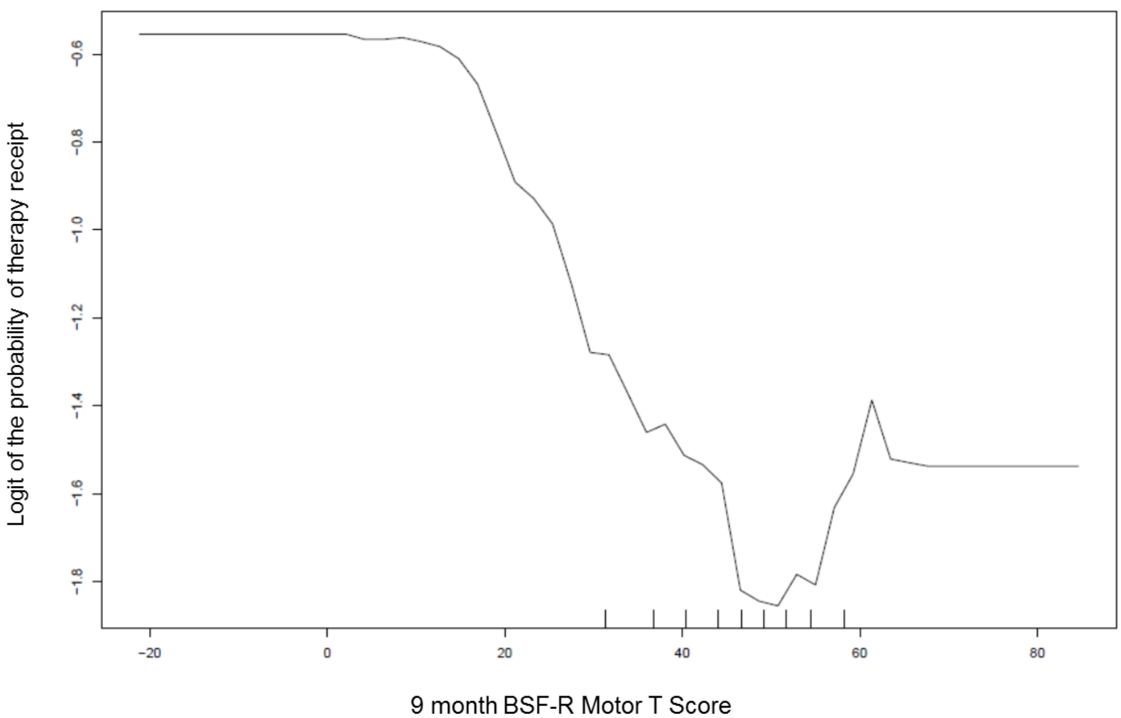
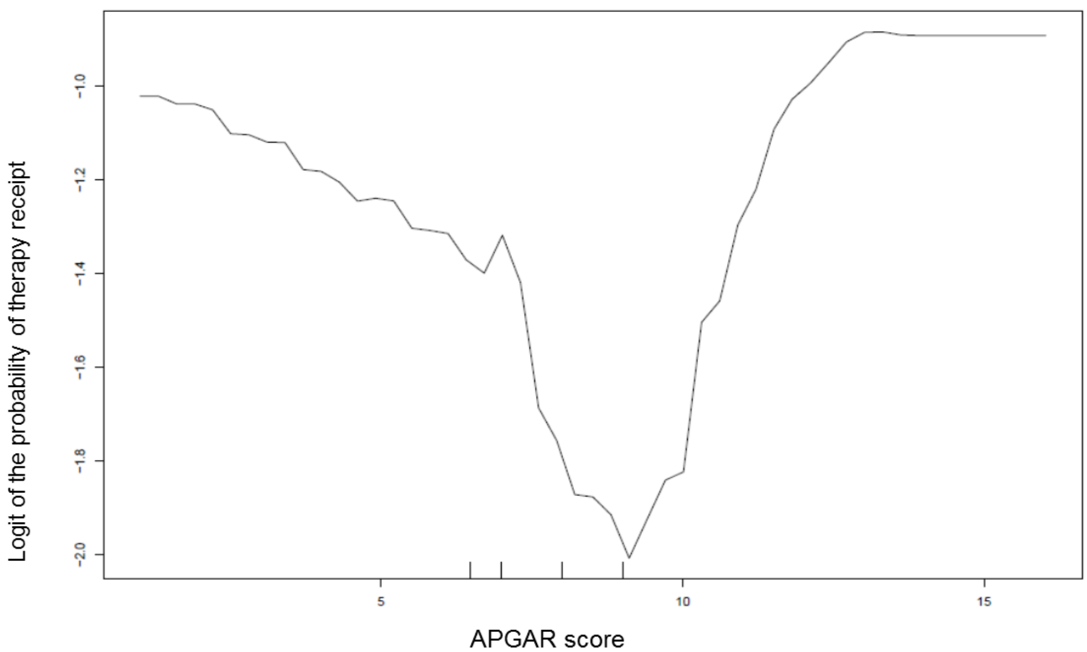


Figure 16. Partial Dependence plot between APGAR score and receiving interventional physical or occupational therapy among low birth weight children



Upper range of APGAR score marginally out of clinical range due to imputed data

Figure 17. Partial Dependence plot between SES and receiving interventional physical or occupational therapy among low birth weight children

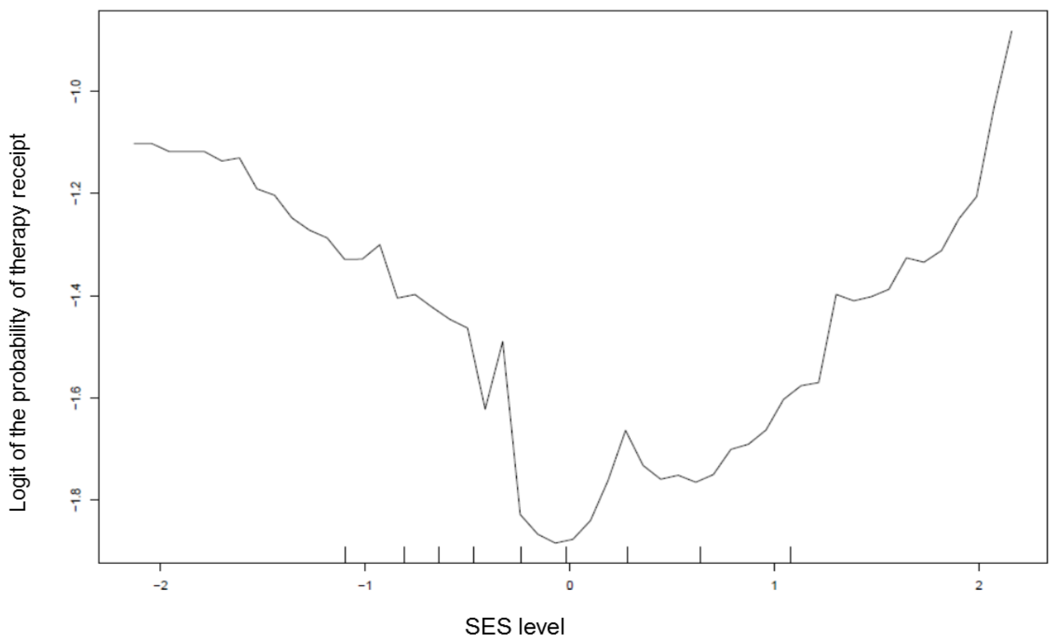


Figure 18. Partial Dependence plot for CESD and receiving interventional physical or occupational therapy among low birth weight children

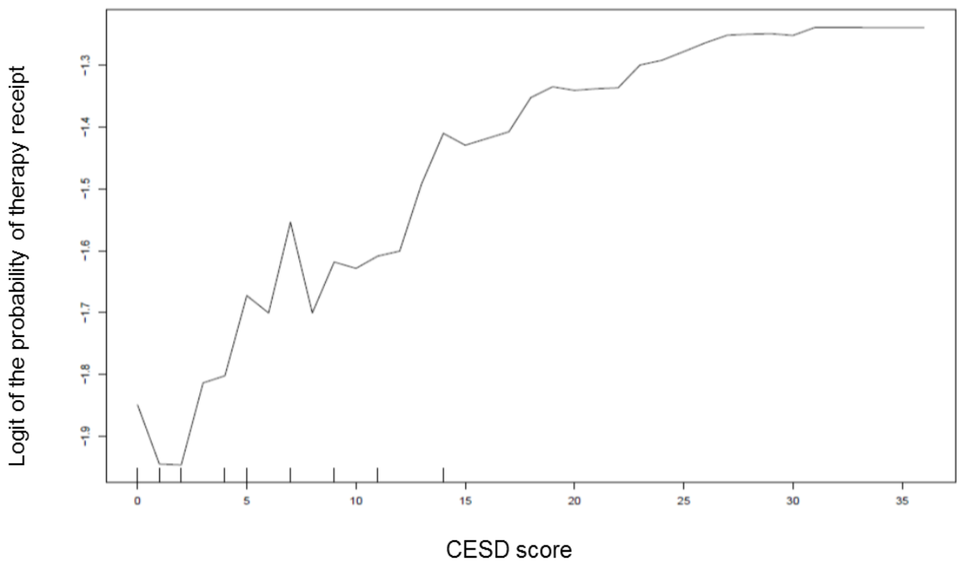


Figure 19. Partial Dependence plot between weekly childcare and physical or occupational therapy receipt among low birth weight children

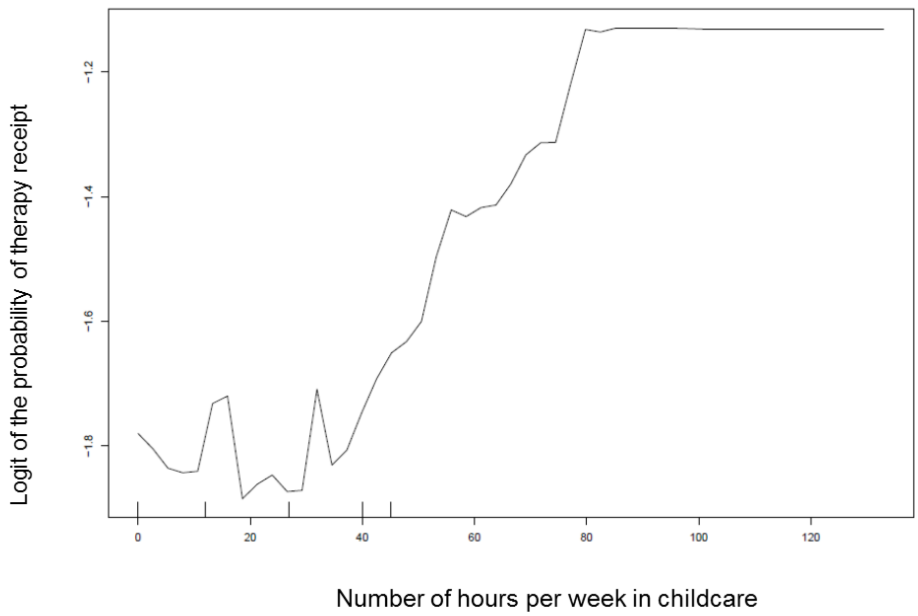


Figure 20. Partial Dependence plot between weekly KIDI score and physical or occupational therapy receipt among low birth weight children

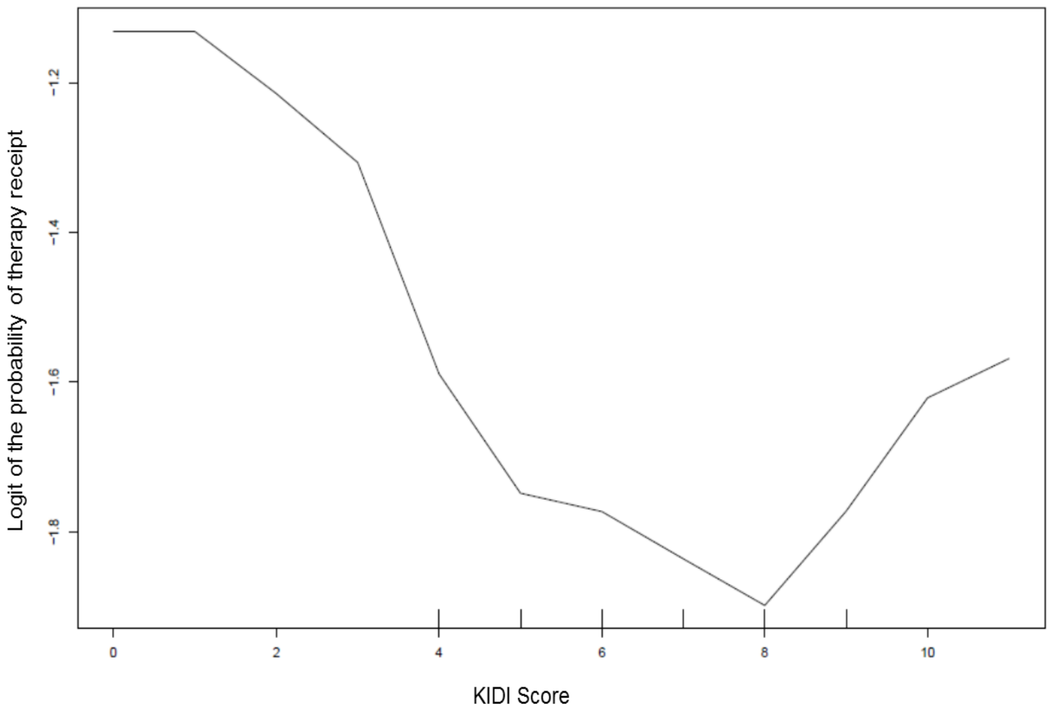
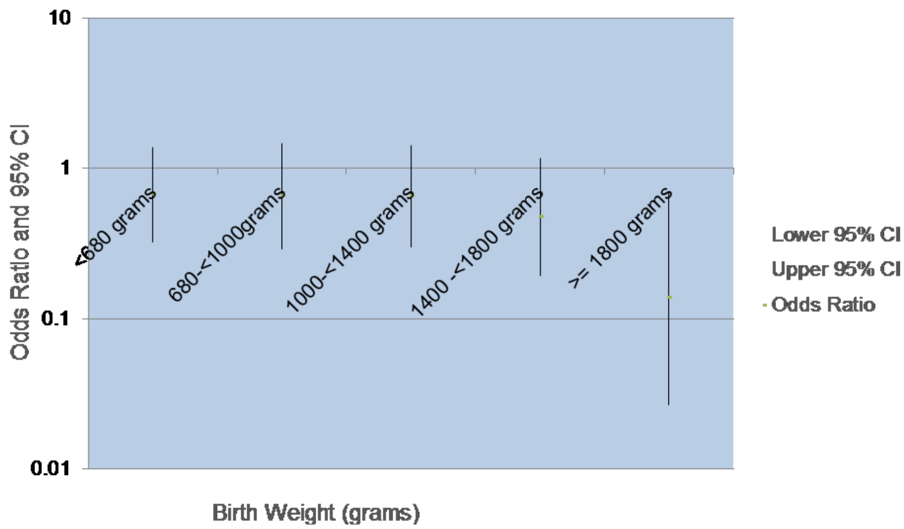
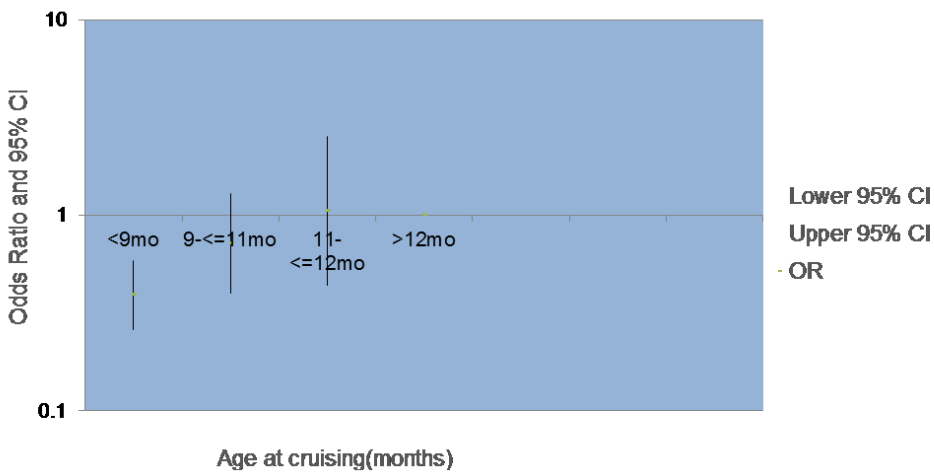


Figure 21. Odds ratios and 95% CI: Birth weight and motor coordination impairment*



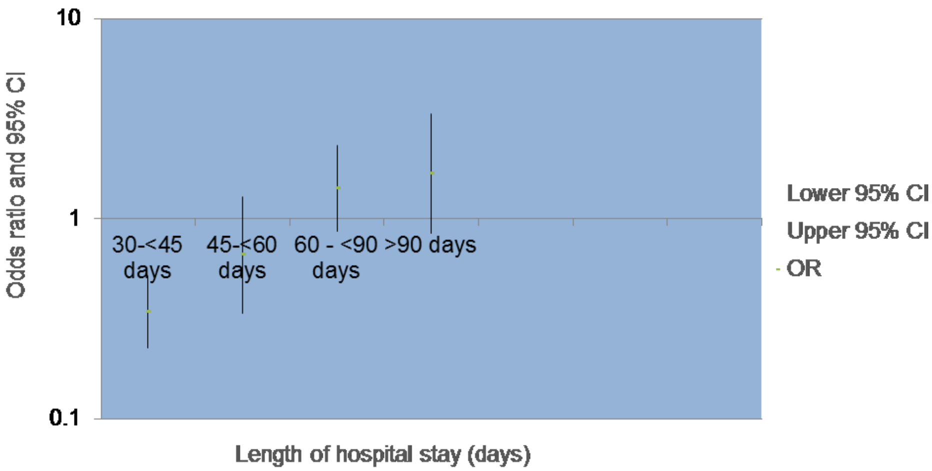
*Analysis among low birth weight children with similar functional ability

Figure 22. Odds ratios and 95% CI: Cruising and motor coordination impairment*



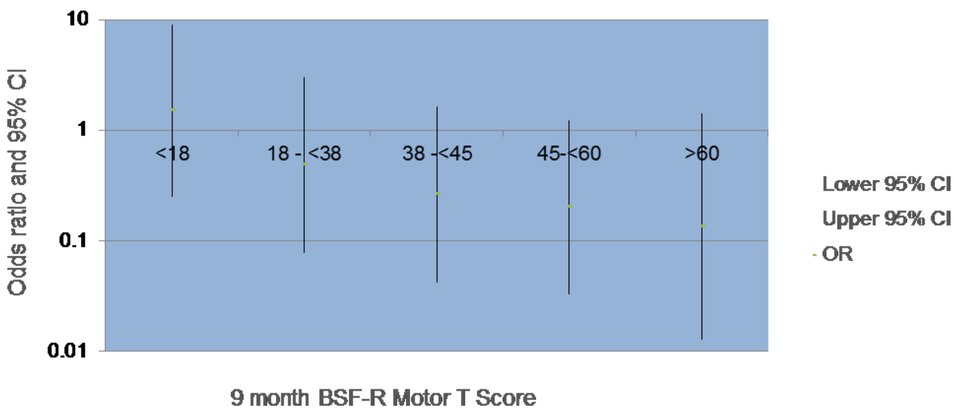
*Analysis among low birth weight children with similar functional ability

Figure 23. Odds ratios and 95% CI: Length of child's hospital stay after birth and motor coordination impairment*



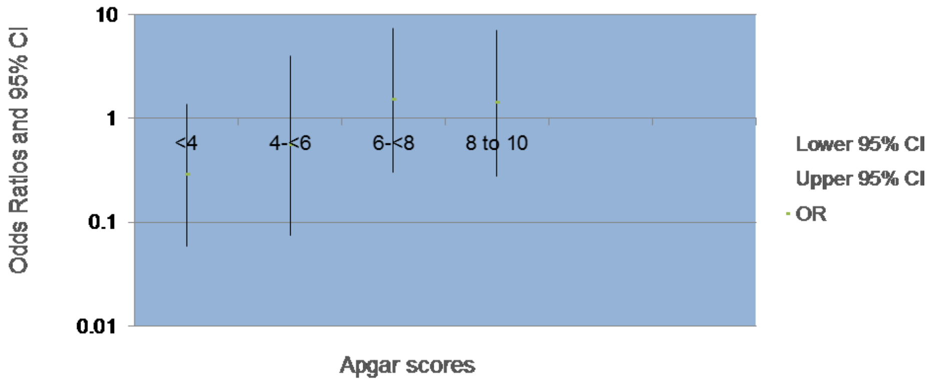
*Analysis among low birth weight children with similar functional ability

Figure 24. Odds ratios and 95% CI: 9 month BSF-R Motor T score and motor coordination impairment*



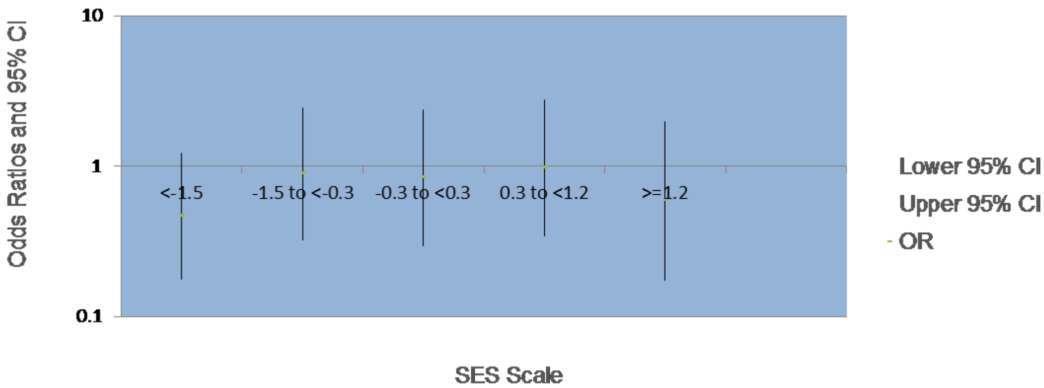
*Analysis among children with low birth weight with similar functional ability

Figure 25. Odds ratios and 95% CI: APGAR score and motor coordination impairment*



*Analysis among low birth weight children with similar functional ability

Figure 26. Odds ratios and 95% CI: SES scale and motor coordination impairment*



*Analysis among low birth weight children with similar functional ability

Figure 27. Directed Acyclic Graph Pathways

Therapy between 9 months and age 2 and preschool motor skills

Therapy ← race/ethnicity → BW → [age of motor milestone] → motor

Therapy ← [BW] → child nutrition → motor

Therapy ← maternal education → hrs in childcare → [age of motor milestone] → motor

Therapy ← maternal education → [income] → hrs in wkly childcare → age of motor milestone → motor

Therapy ← [income] → child nutrition → motor

Therapy ← [income] → Home stimulation → motor

Therapy ← [visual/hearing] → motor

Therapy ← visual/hearing → [age of motor milestone] → motor

Therapy ← visual/hearing → [days in hospital] → motor

Therapy ← Other health needs → [age of motor milestone] → motor

Therapy ← Other health needs → [days in hospital] → motor

Therapy ← [Other health need] → motor

Therapy ← congenital anomaly → [age at motor milestone] → motor

Therapy ← congenital anomaly → [days in hospital] → motor

Therapy <-- time in childcare --> [age at motor milestone] --> motor

Therapy <-- [time in childcare] --> environmental stimulation --> motor

Therapy <-- [Birth weight] --> Motor

Therapy <-- Birth weight --> [Hospital Stay] --> Motor

Therapy <-- Birth Weight --> [Age of motor milestones] --> motor

Therapy <-- Birth Weight --> neurological insult --> [Hospital Stay] --> motor

Therapy <-- [age at motor milestone attainment] --> motor

Therapy <-- Birth Weight <-- GA --> [Age at motor milestone attainment] --> Motor

Therapy <-- Parenting Support --> [Home Env Stimulation] --> Motor

Therapy <-- Income --> [Stimulation in Home] --> Motor

Therapy <-- Negative Pregnancy Behaviors --> [Days in hospital] --> Motor

Community Level Factors --> Therapy --> Motor Development

Therapy <-- [Birth Weight] <-- GA --> Motor

Therapy <-- Congenital Anomaly --> Motor

**Exclude these kids

Therapy <-- [GA] --> Motor

Therapy <-- [Income] <-- Race/Ethnicity --> GA --> Motor

Therapy ← Maternal Education ← Maternal Age → Adverse Neonatal [Days in Hospital] → Motor D

Therapy ← Maternal Education → [Home Environ] → Motor Development

Therapy ← Birth Weight ← [GA] ← Neurological Insult → Motor

Therapy ← [Maternal Education] → Nutrition → Motor Development

Therapy ← [Race/ethnicity] → Birth Weight → Motor Development

Therapy ← [Income] → Time in Childcare → Motor Development

Therapy ← Caregiver health → [Income] → Nutrition → Motor

Therapy ← Caregiver health → [Income] → Home environ → Motor

Therapy ← Caregiver Health → [Income] → Work schedule → Home Envir → Motor

Therapy ← Caregiver Health → Income → Work schedule → Child Nutrition → Motor

Therapy ← Income → Maternal Work Schedule → Child Sleeping Patterns → Motor

Therapy ← [Health Insurance (Medicaid proxy for disease severity)] → Motor

Therapy ← [Failure to Thrive] ← Nutrition → Motor

Therapy ← [Failure to Thrive] → Motor

Therapy ← [Failure to Thrive] ← Income → Motor

Therapy ← [9 month motor score] → Motor

Therapy ← Birth Weight ← Congenital Anomaly → 9 month motor score → Motor

Therapy ← Birth Weight → Neurological Insult → [9 month motor score] →
Motor

Therapy ← Health Insurance ← [Income] ← Single Parent → Home Envir →
Motor

<u>Minimally Sufficient Conditioning Set</u>

GA: Birth Certificate

Days in the hospital after birth: unmeasured covariates are most likely correlated with time in hospital after birth

Age of early milestone attainment

Maternal/Parental Education

Race/Ethnicity: composite variable

Income: baseline

Visual/Hearing deficit: 9 month

Other special health need: 9 month

Birth Weight

Time in childcare

Stimulation in the home: knowledge of child development, HOME score

Failure to Thrive

9 month motor score

REFERENCES

1. Low and Very Low Birthweight Infants. In: Child Trends; 2012.

2. Schendel D, Bhasin TK. Birth weight and gestational age characteristics of children with autism, including a comparison with other developmental disabilities. Pediatrics 2008;121:1155-64.

3. Stephens BE, Vohr BR. Neurodevelopmental outcome of the premature infant. Pediatr Clin North Am 2009;56:631-46, Table of Contents.

4. Developmental Coodination Disorder. Centre for Childhood Disability Research,McMaster University, 2012. (Accessed June 7 2012, 2012, at http://dcd.canchild.ca/en/?_mid_=3276.)

5. Cairney J, Hay JA, Veldhuizen S, Missiuna C, Faught BE. Developmental coordination disorder, sex, and activity deficit over time: a longitudinal analysis of participation trajectories in children with and without coordination difficulties. Dev Med Child Neurol 2010;52:e67-72.

6. Cairney J, Hay JA, Faught BE, Hawes R. Developmental coordination disorder and overweight and obesity in children aged 9-14 y. Int J Obes (Lond) 2005;29:369-72.

7. Cairney J, Hay JA, Faught BE, Wade TJ, Corna L, Flouris A. Developmental coordination disorder, generalized self-efficacy toward physical activity, and participation in organized and free play activities. J Pediatr 2005;147:515-20.

8. Burns YR, Danks M, O'Callaghan MJ, et al. Motor coordination difficulties and physical fitness of extremely-low-birthweight children. Dev Med Child Neurol 2009;51:136-42.

9. Missiuna C, Moll S, King S, King G, Law M. A trajectory of troubles: parents' impressions of the impact of developmental coordination disorder. Phys Occup Ther Pediatr 2007;27:81-101.

10. Edwards J, Berube M, Erlandson K, et al. Developmental coordination disorder in school-aged children born very preterm and/or at very low birth weight: a systematic review. J Dev Behav Pediatr 2011;32:678-87.

11. Salokorpi T, Rautio T, Kajantie E, Von Wendt L. Is early occupational therapy in extremely preterm infants of benefit in the long run? Pediatr Rehabil 2002;5:91-8.

12. Rothberg AD, Goodman M, Jacklin LA, Cooper PA. Six-year follow-up of early physiotherapy intervention in very low birth weight infants. Pediatrics 1991;88:547-52.

13. Riethmuller AM, Jones R, Okely AD. Efficacy of interventions to improve motor development in young children: a systematic review. Pediatrics 2009;124:e782-92.

14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

15. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-81.

16. Drake C. Effects of misspecification of the propensity score on estimators of treatment effect. Biometrics 1993;49:1231-6.

17. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med 2010;29:337-46.

18. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. Pharmacoepidemiol Drug Saf 2008;17:546-55.

19. Martin JA. Births: Final Data for 2005; 2007.

20. McCormick MC. The contribution of low birth weight to infant mortality and childhood morbidity. N Engl J Med 1985;312:82-90.

21. McCormick MC, Brooks-Gunn J, Workman-Daniels K, Turner J, Peckham GJ. The health and developmental status of very low-birth-weight children at school age. JAMA 1992;267:2204-8.

22. Huppi PS, Warfield S, Kikinis R, et al. Quantitative magnetic resonance imaging of brain development in premature and mature newborns. Ann Neurol 1998;43:224-35.

23. de Groot L. Posture and motility in preterm infants. Dev Med Child Neurol 2000;42:65-8.

24. Individual with Diasbilities Education Act. In. Washington DC: One Hundred Fifth Congress of the United States of America; 1997.

25. Feder KP, Majnemer A, Bourbonnais D, Platt R, Blayney M, Synnes A. Handwriting performance in preterm children compared with term peers at age 6 to 7 years. Dev Med Child Neurol 2005;47:163-70.

26. Piek JP, Dawson L, Smith LM, Gasson N. The role of early fine and gross motor development on later motor and cognitive ability. Hum Mov Sci 2008;27:668-81.

27. Diamond A. Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. Child Dev 2000;71:44-56.

28. Losch H, Dammann O. Impact of motor skills on cognitive test results in very-low-birthweight children. J Child Neurol 2004;19:318-22.

29. Drummond PM, Colver AF. Analysis by gestational age of cerebral palsy in singleton births in north-east England 1970-94. Paediatr Perinat Epidemiol 2002;16:172-80.

30. Robertson CM, Watt MJ, Yasui Y. Changes in the prevalence of cerebral palsy for children born very prematurely within a population-based program over 30 years. JAMA 2007;297:2733-40.

31. Vincer MJ, Allen AC, Joseph KS, Stinson DA, Scott H, Wood E. Increasing prevalence of cerebral palsy among very preterm infants: a population-based study. Pediatrics 2006;118:e1621-6.

32. Marlow N, Roberts BL, Cooke RW. Motor skills in extremely low birthweight children at the age of 6 years. Arch Dis Child 1989;64:839-47.

33. de Kieviet JF, Piek JP, Aarnoudse-Moens CS, Oosterlaan J. Motor development in very preterm and very low-birth-weight children from birth to adolescence: a meta-analysis. JAMA 2009;302:2235-42.

34. Bracewell M, Marlow N. Patterns of motor disability in very preterm children. Ment Retard Dev Disabil Res Rev 2002;8:241-8.

35. Marlow N, Hennessy EM, Bracewell MA, Wolke D. Motor and executive function at 6 years of age after extremely preterm birth. Pediatrics 2007;120:793-804.

36. Association AP. Diagnostic and statistical manual of mental disorders. In. 4th ed. Washington, DC: American psychiatric association; 2000.

37. Cantell M, Smyth M, Ahonen T. Clumsiness in adolescence: Educational, motor, and social outcomes of motor delay detected at 5 years. Adapted Physical Activity Quarterly 1994;11:115-29.

38. Bonnier C. Evaluation of early stimulation programs for enhancing brain development. Acta Paediatr 2008;97:853-8.

39. Piek JP. Infant Motor Development. Champaign, IL: Human Kinetics; 2006.

40. Gould E, Tanapat P, Hastings NB, Shors TJ. Neurogenesis in adulthood: a possible role in learning. Trends Cogn Sci 1999;3:186-92.

41. Gould E, Reeves AJ, Graziano MS, Gross CG. Neurogenesis in the neocortex of adult primates. Science 1999;286:548-52.

42. Gould E, Beylin A, Tanapat P, Reeves A, Shors TJ. Learning enhances adult neurogenesis in the hippocampal formation. Nat Neurosci 1999;2:260-5.

43. Shankle WR, Rafii MS, Landing BH, Fallon JH. Approximate doubling of numbers of neurons in postnatal human cerebral cortex and in 35 specific cytoarchitectural areas from birth to 72 months. Pediatr Dev Pathol 1999;2:244-59.

44. Merzenich MM, Recanzone GH, Jenkins WM, Grajski KA. Adaptive mechanisms in cortical networks underlying cortical contributions to learning and nondeclarative memory. Cold Spring Harb Symp Quant Biol 1990;55:873-87.

45. Enhancing the outcomes of low-birth-weight, premature infants. A multisite, randomized trial. The Infant Health and Development Program. JAMA 1990;263:3035-42.

46. Randomised trial of parental support for families with very preterm children. Avon Premature Infant Project. Arch Dis Child Fetal Neonatal Ed 1998;79:F4-11.

47. Hamilton M, Goodway J, Haubenstricker J. Parent-assisted instruction in a motor skill program for at-risk preschool children. Adapted Physical Activity Quarterly 1999;16:415-26.

48. Gianni ML, Picciolini O, Ravasi M, et al. The effects of an early developmental mother-child intervention program on neurodevelopment outcome in very low birth weight infants: a pilot study. Early Hum Dev 2006;82:691-5.

49. Goodman M, Rothberg AD, Houston-McMillan JE, Cooper PA, Cartwright JD, van der Velde MA. Effect of early neurodevelopmental therapy in normal and at-risk survivors of neonatal intensive care. Lancet 1985;2:1327-30.

50. Piper MC, Kunos VI, Willis DM, Mazer BL, Ramsay M, Silver KM. Early physical therapy effects on the high-risk infant: a randomized controlled trial. Pediatrics 1986;78:216-24.

51. Barrera ME, Cunningham CE, Rosenbaum PL. Low birth weight and home intervention strategies: preterm infants. J Dev Behav Pediatr 1986;7:361-6.

52. Girolami GL, Campbell SK. Efficacy of a Neuro-Developmental Treatment Program to Improve Motor Control in Infants Born Prematurely. Pediatric Physical Therapy 1994;6:175.

53. Salokorpi T, Sajaniemi N, Rajantie I, et al. Neurodevelopment until the adjusted age of 2 years in extremely low birth weight infants after early intervention--a case-control study. Pediatr Rehabil 1998;2:157-63.

54. Lekskulchai R, Cole J. Effect of a developmental program on motor performance in infants born preterm. Aust J Physiother 2001;47:169-76.

55. Yigit S, Kerem M, Livanelioglu A, et al. Early physiotherapy intervention in premature infants. Turk J Pediatr 2002;44:224-9.

56. Cameron EC, Maehle V, Reid J. The effects of an early physical therapy intervention for very preterm, very low birth weight infants: a randomized controlled clinical trial. Pediatr Phys Ther 2005;17:107-19.

57. Koldewijn K, van Wassenaer A, Wolf MJ, et al. A neurobehavioral intervention and assessment program in very low birth weight infants: outcome at 24 months. J Pediatr 2010;156:359-65.

58. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol 2010;63:826-33.

59. Breiman L. Statistical modeling: the two cultures. Statistical Science 2001;16:199-215.

60. Provost B, Crowe TK, McClain C. Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales in two-year-old children. Phys Occup Ther Pediatr 2000;20:5-18.

61. The ECLS-B Direct Assessment Choosing the Appropriate Score for Analysis. In: Early Childhood Longitudinal Study Birth Cohort.

62. Yoon DY, Scott K, Hill MN, Levitt NS, Lambert EV. Review of three tests of motor proficiency in children. Percept Mot Skills 2006;102:543-51.

63. Wiart L, Darrah J. Review of four tests of gross motor development. Dev Med Child Neurol 2001;43:279-85.

64. Brown T, Lalor A. The Movement Assessment Battery for Children--Second Edition (MABC-2): a review and critique. Phys Occup Ther Pediatr 2009;29:86-103.

65. VanderWeele TJ, Robins JM. Minimal Sufficient Causation and Directed Acyclic Graphs. The Annals of Statistics 2009;37:1437-65.

66. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol 2006;163:1149-56.

67. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561-70.

68. Strobl C, Malloy J, Tutz G. An introduction to recursive partitioning: rationale,application,and characteristics of classification and regression trees, bagging and random forests: Department of Statistics University of Munich; 2009.

69. Breiman L. Random Forests. Machine Learning 2001;45:5-32.

70. Breiman L. Bagging predictors. Machine Learning 1996;24:123-40.

71. Breiman L. Arcing classifiers. Annals of statistics 1998;26:801-49.

72. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging,boosting, and randomization. Machine Learning 2000;40:1390157.

73. Buhlmann P, Yu B. Analyzing bagging. Annals of statistics 2002;30:927-61.

74. Bylander T. Estimating generalization error on two class datasets using out of bag estimates. Machine Learning 2002;48:287-97.

75. Flury BK, Riedwyl H. Standard distance in univariate and multivariate analysis. The American Statistician 1986;40:249-51.

76. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Stat Med 2006;25:2084-106.

77. Normand ST, Landrum MB, Guadagnoli E, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54:387-98.

78. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg 2007;134:1128-35.

79. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10:585-98.

80. Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. Am J Epidemiol 2010;172:843-54.

81. Deitz JC, Kartin D, Kopp K. Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2). Physical and Occupational Therapy in Pediatrics 2007;4:87-102.

82. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction: Spinger; 2001.

83. Bodnarchuk JL, Eaton WO. Can parent reports be trusted? Validity of daily checklists of gross motor milestone attainment. Journal of Applied Developmental Psychology 2004;25:481-90.

84. Saigal S, Rosenbaum P, Stoskopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. Qual Life Res 2005;14:243-57.

85. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), preschool data collection. In: Education USDo, Statistics NCfE, eds.; 2005-2006.

86. Lipsitt LP. Learning and emotion in infants. Pediatrics 1998;102:1262-7.

87. Gerber RJ, Wilks T, Erdie-Lalena C. Developmental milestones: motor development. Pediatr Rev 2010;31:267-76; quiz 77.

88. Wijnhoven TM, de Onis M, Onyango AW, et al. Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25:S37-45.

89. Wilson BN, Kaplan BJ, Crawford SG, Campbell A, Dewey D. Reliability and validity of a parent questionnaire on childhood motor skills. Am J Occup Ther 2000;54:484-93.

90. Deitz JC, Kartin D, Kopp K. Review of the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2). Phys Occup Ther Pediatr 2007;27:87-102.

91. Liaw A. Classification and regression by randomForest. R News 2002;2:18-22.

92. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics 2005;61:962-73.

93. Barnett LM, van Beurden E, Morgan PJ, Brooks LO, Beard JR. Childhood motor skill proficiency as a predictor of adolescent physical activity. J Adolesc Health 2009;44:252-9.

94. Smits-Engeslman B, Henderson S, Michels C. The assessment of children with DCD in the Netherlands:The relationship between the Movement Assessment Battery for Chidlren and the Korperkoodinations Test fur Kinder. Human Movement Science 1998;17:699-709.

95. Rodger S, Ziviani J, Watter P, Ozanne A, Woodyatt G, Springfield E. Motor and functional skills of children with developmental coordination disorder: a pilot investigation of measurement issues. Hum Mov Sci 2003;22:461-78.

96. Rodger S, Watter P, Marinac J, Woodyatt G, Ziviani J, Ozanne A. Assessment of children with Developmental Coordination Disorder (DCD): Motor, functional, self-efficacy and communication abilities. New Zealand Journal of Physiotherapy 2007;35:99-109.

97. Hack M, Taylor HG, Drotar D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. Pediatrics 2005;116:333-41.

98. Marlow N, Roberts L, Cooke R. Outcome at 8 years for children with birth weights of 1250 g or less. Arch Dis Child 1993;68:286-90.

99. Finken MJ, Inderson A, Van Montfoort N, et al. Lipid profile and carotid intima-media thickness in a prospective cohort of very preterm subjects at age 19 years: effects of early growth and current body composition. Pediatr Res 2006;59:604-9.

100. Schmidt LA, Miskovic V, Boyle MH, Saigal S. Shyness and timidity in young adults who were born at extremely low birth weight. Pediatrics 2008;122:e181-7.

101. Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees: Chapman & Hall; 1984.

102. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008;168:656-64.

103. Segal MR. Machine learning benchmarks and random forest regression (Technical Report). San Francisco , University of California: Center for Bioinformatics and Molecular Biostatistics; 2004.

104. Dunst C, Bruder M, Trivette C, Raab M, McLean M. Natural learning opportunities for infants, toddlers, and preschoolers. Young Exceptional Children 2001;4:18-25.