PREDICTION OF CEREBRAL PALSY IN VERY PRETERM INFANTS

by

Jehier Afifi

A proposal submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

Background: The prevalence of cerebral palsy (CP) is ten times higher in preterm compared to term infants. Accurate and early identification of preterm infants at risk for CP would enable early referral to intervention programs with the potential to improve their functional mobility and quality of life. Large population-based studies of CP in preterm infants have only reported measures of association and did not develop prediction models of CP and assess their diagnostic properties. Furthermore, all these studies used conventional logistic regression for their models. Machine learning may provide more accurate predictions than logistic regression due to its ability to better handle complex relationships between predictors and the outcome. Machine learning methods have not been used yet to predict CP from clinical predictors in former preterm infants.

Objectives: The objective of this study was to develop prediction models for CP in very preterm infants (<31 weeks' gestation) using the random forest (RF) ensemble method and logistic regression and to compare their accuracy in predicting CP.

Study Design: I used a population-based cohort of 777 very preterm survivors from the AC Allen Provincial Perinatal Follow-Up Program Database born between 2000 and 2014 in Nova Scotia. After randomly splitting the sample into training and testing datasets using a 70:30 ratio, clinical and demographic data from the infants and their mothers were used to develop prediction models of CP at three time points (prenatal, perinatal, and postnatal) in the training dataset using RF and logistic regression. Both models were then compared with regard to their discriminative ability (AUC) in the testing dataset.

Results: In this cohort, 86 infants (11%) developed CP. Predictive performance of the models at the prenatal and perinatal time points was poor, regardless of the method used. At the postnatal time point, both RF and logistic regression provided good discrimination of children with and without CP (AUC 0.84 [95% CI 0.74, 0.94] and AUC 0.81 [95% CI 0.74, 0.95], respectively).

Conclusion: Using clinical predictors, logistic regression was comparable to the RF ensemble method in prediction of CP in a population-based cohort of very preterm children. Both methods can be used for predicting CP in former very preterm infants at the time of discharge.

LIST OF ABBREVIATIONS USED

AUC	Area under the curve
BPD	Bronchopulmonary dysplasia
BSITD	Bayley Scale of Infant and Toddler Development
CART/CT	Classification (and regression) tree/classification tree
CP	Cerebral palsy
CPAP	Continuous positive airway pressure
GMFCS	Gross motor functional classification system
IVH	Intraventricular hemorrhage
LR	Likelihood ratio
NDI	Neurodevelopmental impairment
NEC	Necrotizing enterocolitis
NICU	Neonatal intensive care unit
NPV	Negative predictive value
OR	Odds ratio
PFUP	Perinatal Follow-Up Program
PPV	Positive predictive value
PVL	Periventricular leukomalacia
RF	Random forest
ROC	Receiver operator characteristics curve
ROP	Retinopathy of prematurity
RR	Relative risk
SGA	Small for gestational age
VLBW	Very low birth weight

Statement

Preterm infants are at risk of adverse neurodevelopmental outcomes including cerebral palsy (CP). Identification of preterm infants who are at high risk for CP, through prediction models, may facilitate early referral to intervention programs, with the potential to improve their mobility and their quality of life. Traditionally, logistic regression has been used to develop prediction models for CP in literature. The random forest (RF) ensemble method is a statistical machine learning method that has been shown to provide accurate prediction and may improve the overall performance of the prediction model when compared to logistic regression, particularly with large multidimensional data. The advantages of RF, over logistic regression, include its ability to accommodate a large number of predictors when the sample size is small and to handle complex (non-linear) relations between the predictors and the outcome, in addition to not relying on assumptions about the distribution of the predictors or the outcome variables. To the best of my knowledge, the RF method has not been used to predict CP in population-based studies of preterm infants, using only clinical predictors that can be readily abstracted from patient records. The study used the AC Allen Provincial Perinatal Follow-Up Program Database to develop prediction models of CP using RF and logistic regression. Both models were compared with regard to their accuracy in predicting CP among very preterm infants.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Stefan Kuhle *MD PhD,* Perinatal Epidemiology Research Unit at the IWK Health Centre, for his guidance and constant supervision. He introduced me to the area of machine learning in risk prediction and secondary use of data. His door was always open whenever I have a question about my thesis or statistical analysis. His mentorship helped me throughout the learning process of this project. I really appreciate his patience and understanding of my professional duties and busy schedule.

I would like to thank all my committee members for their guidance and support. I am deeply grateful to Christy Woolcott *PhD*, Perinatal Epidemiology Research Unit at the IWK Health Centre, for her insightful input, review and encouragement that guided me through the writing of this thesis. I am thankful for Kathy MacPherson *MD MPH*, Associate Professor and former Graduate Coordinator at the Department of Community Health and Epidemiology and Michael Vincer *MD*, Associate Professor of Pediatrics at Dalhousie University, Neonatologist and Director of the AC Allen Provincial Perinatal Follow up Program Database at IWK Health Centre, for their continuous support and valuable review of this thesis.

I am gratefully indebted to my colleagues at the Division of Neonatology, at IWK Health Centre, for their help and support. I must express my special gratitude and thanks to my beloved family for their continuous encouragement and support that helped me achieve my goals.

Finally, I dedicate this work to the memory of my parents and to the newborn infants and their families that I care for throughout my career.

CHAPTER 1: INTRODUCTION

Survival of preterm infants has increased over the past several decades (1-3). However, the rates of neurodevelopmental impairment (NDI), including CP, remain high (4-7). Cerebral palsy is defined by a heterogeneous group of clinical signs describing permanent disorders of movement and posture, with half of the affected children having either limited or no walking ability (8). The disease is commonly associated with other comorbidities (seizures, cognitive and behavioral disorders) which require multidisciplinary health services utilization (9). The economic burden of prematurity in Canada estimated a national cost of $ 587.1 million for all preterm infants with the cost per infant over the first 10 years of life being different based on gestational age; $67,467 per infant for those < 28 weeks' and $52,796 per infant for those 28-32 weeks' (10).

The rate of CP is 10 times higher in preterm and very low birth weight infants (VLBW) compared to term infants (8). Accurate and early identification of preterm infants at risk for CP would enable early referral to intervention programs with the potential to improve their functional mobility and quality of life (11). The majority of studies that examined the risk factors associated with development of CP, including large population-based studies, reported measures of association (i.e., odds ratio (OR), relative risk (RR)) and not diagnostic properties of prediction (12-21). Additionally, studies that reported the quantitative measures of prediction of CP (sensitivity, specificity, positive and negative likelihood ratios (LR)) were limited by small sample size and their use of convenience samples (List of studies in Appendix 1). Traditionally, all these studies used logistic regression.

For binary outcomes such as CP, random forest (RF) is an ensemble of classification trees (CT), each constructed in a bootstrapped sample with a random subset of possible predictors (22). Each CT is built recursively by successive divisions or binary splits that maximize the discrimination of those who developed the outcome of interest from those who did not at each split (22). The RF then votes for the optimal classification or the majority vote. Using randomness in building each tree in the forest leads to a better prediction and does not have the problem of overfitting (23). The RF is likely superior to logistic regression as it accommodates a large number of predictors relative to the observations, it considers non-linear relations between predictors and the outcome or high order interaction between various predictors, and it does not require assumptions of distribution of predictors (23). Importantly, this method has not been used in the context of predicting CP in preterm born children from clinical predictors to date.

CHAPTER 2: OBJECTIVES

My study's primary aim was to develop a prediction model of CP in very preterm infants (defined as those who were born before 31 weeks gestation) using the AC Allen Provincial Perinatal Follow Up Program (PFUP) database. The following specific objectives were examined:

Objective 1: To develop a prediction model of CP in very preterm infants using logistic regression and describe the predictors of CP in this patient population.

Objective 2: To develop a prediction model of CP in very preterm infants using RF ensemble method.

Objective 3: To compare the models developed, using both logistic regression and RF, with respect to their ability to discriminate between children who do develop CP from those who do not.

CHAPTER 3: BACKGROUND

3.1 Cerebral Palsy in Preterm Infants

3.1.1. Burden of Preterm Birth

Prematurity is defined as infants who are born before complete 37 weeks of gestation and can be further sub-categorized into: (i) extremely preterm infants (< 28 weeks gestation), (ii) early preterm infants (28^0 weeks to 31^6 weeks gestation) and (iii) late preterm infants (32^0 weeks to 36^6 weeks' gestation) (3). Preterm birth and its consequences constitute a major health problem in Canada and worldwide. In Canada, the rate of preterm births increased from 7.0% in 1995 to 7.8% in 2013 (1). In Nova Scotia, infants born very preterm (22^0 weeks to 30^6 weeks gestation) constitute 1% of the total births and around 10% of the annual admissions to the Neonatal Intensive Care Unit (NICU) (5).

Prematurity has significant societal impact due to the considerable emotional burden and economic costs to families of preterm children and the increased health service utilization among preterm survivors (10). This patient population represents a subgroup of the community with a wide range of health needs requiring multiple resources to provide the necessary medical, developmental, educational, and family support.

Advances in perinatal and neonatal care have led to improved survival of preterm infants over the past several decades (2); however, the rates of CP and NDI (defined as any of the following: CP, cognitive delay, language delay, visual or hearing impairment), remain high (4-7). In addition to NDI, preterm and VLBW (whose birthweight is less than 1500 grams) infants are at high risk for learning disability, and behavioral disorders such as attention deficit hyperactivity disorder or autism (4, 7). There is an inverse relation between the gestational age and intact survival (survival without major NDI), with the effect of gestational age being particularly strong in extremely preterm and VLBW infants (4, 7).

3.1.2. Cerebral Palsy

Cerebral palsy is defined as a non-progressive developmental disorder affecting muscle tone, movement and posture and causing mobility restriction or disability that originates from insults affecting the fetal or infant brain (8, 24, 25). In addition to motor impairment, CP is commonly accompanied by other comorbidities such as seizures, sensory impairment, cognitive delay and communication or behavior disorders (9, 26). The impact on children's lives and their families continues through adolescence and into adulthood.

No single test is available to rule in or rule out CP, and the diagnosis is entirely based on clinical neuromotor assessment of muscle tone, posture and movement. Cerebral palsy is classified into different subtypes based upon the underlying abnormalities of the muscle tone, the anatomical distribution and the severity of motor impairment (8, 24). Based on the abnormality of the muscle tone, CP is divided into the following groups: spastic (commonest), athetotic, hypotonic or mixed types. Spastic CP is further classified into four subtypes, based upon the distribution: monoplegia (if only one arm or one leg is affected); hemiplegia (if one arm and one leg on the same side are affected, asymmetric spasticity), diplegia (both legs are more affected than arms, symmetric spasticity) or quadriplegia (both arms and both legs are affected equally, bilateral symmetric spasticity) (8, 24, 25). Finally, once CP diagnosis is confirmed, grading of CP severity is conducted based on ambulation with or without aids.

The mean age at diagnosis of CP is around 12-18 months of corrected age (defined as the chronological age in weeks minus the number of weeks a preterm infant is born before complete 40 weeks of gestation). By 5 years of age, majority of children with CP have established comorbidities and spasticity that greatly impacts their quality of life. Many of these comorbidities are modifiable, if early identification and referral to appropriate services was initiated using the window of brain plasticity with the potential to optimize their motor and cognitiveoutcomes, prevent secondary complications, and importantly empowers and enhance the well-being of their caregivers (27). On the other hand, because subtle or mild CP can be diagnosed as late as 24-36 months or even later, serial neuromotor assessment of children who were born preterm is necessary (26).

3.1.3. Preterm Infants Are at High Risk for Developing Cerebral Palsy

Cerebral palsy is the most common neuromotor disability in children, with a reported prevalence of 1.5-2.5 per 1000 live births (8). Cerebral palsy is 10 times more common in preterm and VLBW infants compared to term infants, with nearly half of cases being former preterm infants (3, 28). Cerebral palsy was reported in 44 per 1000 live births among children born <32 weeks gestation and in 60 per 1000 live births among VLBW children (28). It is controversial whether the high incidence rates of CP among preterm infants is explained by the increased survival of very preterm infants over the last few decades (28, 29). A population-based study in Nova Scotia (1988-2007) showed that 10.6% of infants born at a gestational age $22^{0/7}$-$30^{6/7}$ weeks who survived for at least one year developed CP, with 42% of these children having either limited or no walking ability (29). The study also reported increased prevalence of CP over a 20-year study period, from 5.5 per 10,000 live births during the first epoch (1988-1992) to 9.2 per 10,000 live births during the third epoch (1998-2002), that was not attributable to increased survival (29).

Cerebral palsy is one of the most devastating consequences of preterm birth. Preterm birth is commonly associated with exposure to risk factors, such as hypoxia/ischemia or infection that induce a picture of encephalopathy of prematurity (30). This condition is characterized by brain injury or dysmaturation/disruption of the normal developmental trajectory and growth of neuronal cells or both. Oligodendrocytes, microglia and astrocytes have a crucial role in brain development and microstructural connections of neuronal pathways (31, 32). They are also highly vulnerable having high affinity to calcium influx and over-expression of glutamate receptors. Activation of these cells in the fetal or preterm brain results in a cascade of biochemical responses and release of cytotoxic mediators (such as free oxygen radicals) at a time of critical brain development (31, 32). These mediators damage the pre-myelinated white matter axons of the developing brain of preterm infants and induce injury characterized by destruction and apoptosis (31, 32). Brain disruption is characterized by failed maturation of oligodendrocytes to myelinating oligodendrocytes or abnormal organization of cortical neurones leading to myelination abnormalities and reduced brain volumes in children born preterm (33, 34).

3.1.4. Risk Factors for CP

Various factors have been identified as potential risk factors for CP in preterm infants. These risk factors are interrelated and are inversely related to gestational age, with extreme preterm infants being at the highest risk, due to the vulnerability of the brain cells (30). The general hypothesis is that these risk factors are associated with injury to the developing fetal or neonatal brain, as outlined above, antecedent to CP.

Potential risk factors for CP in preterm infants may be largely classified into prenatal, perinatal, and postnatal factors. Prenatal factors can be either maternal (chorioamnionitis) (35-38) or fetal factors (fetal growth restriction, male sex) (39-45). Perinatal factors include low gestational age at birth and birth depression (46-50). Postnatal factors include severe neurologic injury (defined as severe intraventricular hemorrhage (IVH) (51, 52) or cystic periventricular leukomalacia (PVL) (53-55), sepsis (56-58), necrotizing enterocolitis (NEC) (59, 60), bronchopulmonary dysplasia (BPD) (61, 62), postnatal steroids (63-68) and neurosensory impairment such as retinopathy of prematurity (ROP) (69-73).

1. Chorioamnionitis is defined as inflammation of the placental membranes and has been reported to increase the risk of CP in term infants by 2- to 12-fold (35). Both histologic and clinical chorioamnionitis increase the risk of neonatal morbidity and NDI, and they are more often associated with diplegia than with other subtypes of CP (35). The risk is even higher in preterm infants, as preterm birth is often thought to be secondary to in-utero infection or chorioamnionitis (36). This fetal inflammatory status predisposes to preterm labor, fetal white matter brain injury and chronic lung disease, all of which are antecedents of CP (36-38). The fetal white matter insult in preterm infants, induced by cytokines inflammatory mediators, is identified clinically after birth as cystic PVL and subsequent CP (37).

2. **Intrauterine growth restriction** is defined as a suboptimal uterine environment and placental insufficiency that affects both fetal body and brain growth rates. Utero-placental insufficiency eventually results in infants being born as small for gestational age (SGA), defined as birth weight < 10^{th} percentile for gestational age and sex. Preterm infants who are SGA, have higher risk of CP (44% vs 6%; OR 11., 95% CI 6.25-22.08) compared to those who are not SGA (39). Even among children with CP, being SGA significantly increased the risk of NDI regardless of gestational age; data from the Canadian Registry of CP showed that children with CP who were SGA had significantly higher impairment of the fine motor (RR 1.46, 95% CI 1.02-2.11), gross motor (RR 1.53, 95% CI 1.12-2.10), language (RR 1.24, 95% CI 1.10-1.40), and cognitive development (RR 1.33, 95% CI 1.06-1.69) when compared to children with CP who were not SGA (40). The association of SGA with CP was attributed to prenatal risk factors associated with SGA (such as utero-placental insufficiency, genetic anomalies, fetal infection) that may contribute to brain injury antecedent to CP (40-42).

3. **Male** preterm infants have higher mortality, morbidity and NDI compared to female preterm infants (43). A large population based study of very preterm infants, born before 33 weeks' gestion, reported that male sex was independently associated with CP at 5 years of corrected age, after controlling for cerebral injury and obstetric risk factors (OR 1.52; 95% CI 1.03–2.25) (21). Another large population-based study of extremely preterm infants, born before 27 weeks' gestation, showed that boys had significantly lower mean composite cognitive and language scores compared to girls (43). Male sex has been known to be associated with high rates of severe respiratory insufficiency (severe respiratory distress syndrome) and associated comorbidities (postnatal steroids, BPD and abnormal brain imaging) that may increase the risk of CP in these infants (44). The higher rates and severity of respiratory disease in preterm boys than girls was hypothesized to be due to in-utero exposure to Mullerian inhibiting factor and androgen, leading to decreased surfactant production (45). However, Peacock et al. reported that male sex remained significantly associated with NDI in preterm infants even after adjustment for gestational age, birth weight, BPD, and abnormal neuroimaging, suggesting an intrinsic male effect to be contributing to the poorer outcomes in preterm boys (44).

4. **Low gestational age** is commonly associated with exposure to hypoxia/ischemia or infection-related events that can induce fetal brain injury at a critical time of brain growth and development and may interfere with establishing the complex microstructural connections throughout the fetal brain (30, 31, 46). One of the theories underlying spontaneous preterm birth is thought to be secondary to an intrauterine infectious process that induces intra-amniotic inflammatory response with activation of cytokines and chemokines that precipitates premature uterine contractions (47). The fetal brain growth during the third trimester of pregnancy results in a four- to five-fold increase in the brain volume including the cortical grey matter, white matter and cerebellum (48). This fetal brain growth is also accompanied by complex brain development at the cellular level including; neuronal migration, proliferation and myelination (46, 48). Therefore, low gestational age infants, born before the third trimester, have small-sized brains with simple primitive structure that make them vulnerable to develop brain injury and subsequent NDI, including CP. Gestational age has been shown to have an inverse relationship with CP in general, particularly spastic diplegia (30).

5. **Birth depression or intrapartum asphyxia** involves multiple perinatal factors that ultimately result in reduction of blood flow or oxygen delivery to the fetal brain manifested as a need for cardiopulmonary resuscitation at birth, low Apgar scores, and fetal/neonatal hypoxia and acidosis, collectively known as birth depression. The condition is associated with increased mortality and adverse outcomes in both preterm and term infants (49). Cerebral palsy was believed to occur secondary to intrapartum asphyxia, but recent reports showed that birth depression may account for only 10% of cases of CP and that the timing of CP may be related to antepartum, intrapartum or even postpartum hypoxic/ischemic events (49). However, the association of birth asphyxia with CP is debatable in preterm and VLBW infants, where the majority receive resuscitation at birth, have low Apgar scores and evidence of hypoxia and acidosis related to prematurity and respiratory insufficiency rather than birth asphyxia.

6. **Severe neurologic injury** related to prematurity is associated with NDI and CP (13, 52). Severe neurologic injury in preterm infants is defined as the presence of one or more of the following: severe hemorrhage with ventricular dilatation or parenchymal bleeding (also called grade 3/4 IVH, as per Papile classification) (51), and or cystic PVL). These brain injuries may be clinically silent and can only be identified on cranial imaging. Therefore, routine sequential cranial ultrasound screening has been the standard of care for very preterm infants. A recent meta-analysis showed that parenchymal hemorrhage (with or without ventricular dilatation) in preterm infants was associated with increased risk of CP (RR 3.4, 95% CI 1.60-7.22; 9 studies 2876 infants), whereas both cystic and non-cystic PVL were independently associated with CP (RR 19.12 (95% CI 4.57-79.90) and RR 9.27 (95% CI 5.93-14.50), respectively, 2 studies 802 infants) (52). Spastic diplegia is the most common sequala of PVL in preterm infants (52-54). This is attributed to the anatomical distribution of PVL involving the descending fibers from the motor cortex to the internal capsule in close proximity to the periventricular area. Additionally, PVL involves the neuronal tracts of the visual, auditory, and somatosensory regions, therefore PVL-induced CP in preterm infants is associated with visual and auditory impairment, cognitive or language delay and epilepsy (52-54).The severity of PVL is inversely related to gestational age and birth weight, with quadriplegia being common in severe PVL (equal or more than grade 2) (54). In a large population based French cohort of 1812 infants born before 33 weeks' gestation (EPIPAGE), the prevalence of CP was 61% among infants with cystic PVL, 50% among those with parenchymal haemorrhage, 8% among those with grade I-IVH, and only 4% among those with undetectable cerebral injury on brain imaging (21). Similar findings of increased prevalence of CP in relation to the severity of PVL were also reported by Resic et al. (53). Animal models showed that PVL-induced CP in preterm infants being characterized by white matter necrotic lesions, hypomyelination, microglial activation, astrogliosis and neuronal death, with injury to oligodendrocytes being the first step in PVL (55). The underlying mechanism is thought to be due to hypoxia ischemia with or without infection (55).

7. **Sepsis**, with or without meningitis, is a common morbidity of preterm infants. Both early and late neonatal sepsis are common morbidity of prematurity and are caused by a variety of organisms including group B streptococci, gram negative rods, candida or coagulase-negative staphylococci. Among very preterm infants,Danish study, those with sepsis have 3 times higher odds CP compared to without sepsis (56, 57). The EPIPAGE study (n=1812, < 33 weeks' gestion) examined the effect of sepsis on the neurodevelopment at five years of age, the authors reported that exposure to maternal-fetal infection was associated with 2 times higher odds of CP (OR 2.13 (95% CI 1.28–3.55) (21). Similar findings were reported by a Swiss national cohort of very preterm infants born between 2000 and 2007 (n=541, gestation 24-27 weeks) who had 3 times higher odds of CP among those with proven sepsis (OR 3.23 (95% CI 1.23-8.48) (57). The impact on CP is additive in preterm and VLBW infants, if sepsis and hypoxia/ischemia co-existed (58). Sepsis may be a manifestation of immunodeficiency related to prematurity and the resultant vulnerability of the developing brain. The underlying mechanism of brain injury with sepsis was attributed to either systemic inflammatory response with influx of cytokines or cerebral ischemia/reperfusion secondary to systemic hypotension during bacteremia (30, 31, 58).

8. **Necrotizing enterocolitis** (NEC) is a devastating disease of prematurity resulting in severe gut ischemic necrosis and intestinal failure and is associated with increased mortality, short- and long-term morbidities among survivors. In a Danish study, preterm infants with NEC who were assessed at 36 months of corrected age, had significantly higher odds of CP (OR 1.5, 95% CI 1.2-2.0) compared to those without NEC (59). This detrimental effect persists until school age with associated gut-related morbidities (such as: presence of stoma, prolonged parenteral nutrition, line-associated complications and frequent hospitalization) that significantly impact growth, sensory, motor and cognitive development (60). The underlying mechanisms of brain injury secondary to NEC include concomitant sepsis, release of inflammatory cytokines, hemodynamic instability, and ischemia/reperfusion injury (58, 59).

9. **Bronchopulmonary dysplasia** (BPD), defined as oxygen dependency at 36 weeks postmenstrual age with or without mechanical ventilation, is a common and serious complication of prematurity. The condition is considered to be an inflammatory disease with early injury to the premature lungs and is associated with NDI, impaired executive functions, overall growth and respiratory morbidities that extend up to adulthood (61). Van Marter et al (2011) reported that very preterm infants with severe BPD, requiring mechanical ventilation at 36 weeks postmenstrual age, have a six-fold increased risk of quadriplegic CP and a fourfold increased risk of diplegia (62). The postulated mechanisms for underlying brain injury include inflammatory mediators, frequent episodes of hypoxemia, and systemic steroid therapy (30, 31).

10. **Postnatal systemic steroids** are used to facilitate extubation and to reduce the complications of severe BPD in preterm infants. However, a Cochrane review and a meta-analysis of 26 clinical trials reported that preterm infants who received early postnatal steroids (< 8 days) had significantly higher rates of CP or abnormal neurologic examination compared to controls (63, 64). For late steroids (≥ 8 days), the increased rates of CP were partly offset by a reduction in late mortality and, consequently, the composite outcome of death or CP was not significantly different between the late steroid group and controls (63, 65, 66). The detrimental effect of systemic steroids on brain development is supported by studies in animals and humans. In neonatal animals, pharmacological doses of steroids are associated with impaired brain development at cellular level that led to delayed brain growth and maturation (67). Murphy et al (2001) showed that preterm infants treated with dexamethasone for BPD had 35% lower brain cortical gray matter volume on brain magnetic resonance imaging compared to untreated infants (68).

11. **Severe retinopathy of prematurity** (ROP), defined as > stage 2 ROP or requiring intervention, is the commonest cause of childhood blindness. The disease is characterized by an initial phase of vascular arrest of the developing retina followed by a neovascularization phase with subsequent retinal detachment if untreated. Inflammation, being a control factor for angiogenesis and neovascularization, is a major contributor to the development of ROP; the associated release of cytokines and other inflammatory mediators that extends beyond the visual cortex and the visual pathways to other areas of the developing preterm brain (69, 70). In addition to the cognitive delay secondary to visual impairment, children with severe ROP develop motor delay with abnormal coordination and lower scores of standardized movement assessment compared to those without severe ROP (71-73).

In addition to the above-listed risk factors associated with CP, there are limited data from longitudinal studies of preterm infants on several routinely collected factors that may potentially be associated with CP or antecedents of CP including: maternal age, maternal chronic illnesses, maternal exposure (medications, smoking, illicit drugs or alcohol use), maternal education, socioeconomic status, or being a single parent, among other factors. Generally, the results from these reports were inconsistent.

3.1.5 Tertiary Prevention of CP

Prediction of CP in very preterm infants enables identification and early referral of high-risk infants to intervention programs that have the potential to improve their mobility and cognitive development (11, 27, 74, 75). Traditionally, multiple disciplines have been involved in managing children with CP and offering various interventions, including rehabilitation, physiotherapy, medications (botulinum toxin injections), orthoses and surgical interventions. Early intervention for children with CP has been shown to improve their motor function, their cognitive and language development (27, 74, 75). Early intervention has been also shown to improve hand function in children with hemiplegic CP and ambulation in preterm born children with diplegia (74). Parental involvement and family integration in early intervention programs has been known to improve children's development and behavior, particularly in relation to communication and relationships, independence, and community participation (76, 77). Promoting parenting skills is thought to play a major role, not only by improving the cognitive and behavioral outcomes of children affected by CP, but also by decreasing anxiety and depression of their caregivers (78).

3.2 Prediction of CP in Preterm Born Children

Multiple studies evaluated the association between various risk factors and the development of CP in preterm born children (12-21). The majority of these studies, including large population-based studies, reported measures of association (i.e. OR, RR) and not diagnostic properties of prediction. Additionally, studies that reported the quantitative measures of predictive accuracy (sensitivity, specificity, positive and negative LR) were limited by small sample size and their use of convenience samples.

3.2.1 Population-based Studies of CP in Preterm Infants

Large population-based studies of preterm infants have investigated the role of multiple exposures on the development of CP as the primary outcome or as a component of a composite outcome of NDI (defined as CP, cognitive or language delay, deafness or blindness) (19-21, 29). Notably, all these studies used logistic regression analyses to identify risk factors associated with CP among these large population cohorts.

The EPIPAGE study, a large population-based prospective cohort of 1812 infants born <33 weeks of gestation in 1997 in France, reported CP in 14% of survivors and showed increased rates of NDI with decreasing gestational age (21). The authors primarily examined the role of neuroimaging in predicting CP and showed that significant neuroimaging abnormalities, particularly cystic PVL and parenchymal hemorrhage (formerly called grade 4 IVH), were independently associated with CP at 5 years of age. The EPICure study, a large population-based prospective cohort of 1031 surviving preterm infants (<28 weeks gestation) who were born between 1995-2006 in the United Kingdom and Ireland, showed that CP (14% of survivors) was more prevalent in children with gestational age < 26 weeks compared to those with gestational age of 26-27 weeks (19). The authors reported a trend of improvement in survival without disability over the study period, particularly in extreme preterm children (24-25 weeks gestation) (19). Similar outcomes were reported by Leversen who followed a prospective cohort of 371 extremely preterm infants born before 28 weeks' gestation in Norway and reported CP (alone or combined with NDI) in 11% of survivors (20). In Nova Scotia, Vincer prospectively followed a cohort of 1430 preterm infants < 31 weeks' gestation over a 20-year period (1988-2007) divided into four epochs. The study reported CP in 11% of preterm survivors (106 of the 1106), with peak prevalence in the third epoch (1998-2002) that is not attributed by the increased survival of extremely preterm infants, as the lower mortality rates did not correlate with the prevalence of CP (29). The study also showed that maternal anemia, use of tocolytics during pregnancy and infant's home oxygen therapy were highest during the peak prevalence of CP (29).

3.2.2 Prediction Studies of CP in Preterm Infants

Three reviews explored studies that reported prognostic factors and predictors of CP in high risk term and preterm newborn infants (79-81). A list of studies that reported the diagnostic properties for prediction of CP, alone or as part of NDI, in preterm infants together with the developmental tests used are provided in Appendix 1 (82-111). The selected studies (n=21) were limited to those published between 2000 and 2019 and with birth cohorts starting from 1990, known as the post-surfactant era; as the neonatal mortality and short and long-term morbidities significantly changed after introduction of antenatal steroids and surfactant therapy in management of preterm birth. Studies with a follow up of more than 5 years were not included, as the context of prediction and CP outcomes at this age are not comparable to the goal of this analysis and would yield different predictors. The majority of studies were prospective cohorts (n=14), including two cohorts from randomized clinical trials, and the remaining were retrospective cohorts or case control studies (n=7). The studies originated from eleven developed countries including Canada: Sweden, Norway, Germany, Netherlands, Austria, Italy, Australia, New Zealand, Japan, United States and Canada, whereas two studies were conducted across multiple countries. Seventeen studies reported CP as the primary outcome or separately reported if it is part of a composite outcome, while the remaining studies (n= 4) reported a composite outcome of NDI, including CP. The most common method used to classify CP in the majority of studies was according to the Gross Motor Function Classification System (GMFCS) by Palisano (83) or Hagebrg (87). The gestational age at birth of included infants varied; however, twelve studies reported outcomes in very preterm infants born before 32 weeks' gestation. The predictors of CP in these studies included: amplitude-integrated electroencephalography (n=2) (82, 84), cranial ultrasound (n=3) (85, 86, 88), brain MRI (n=3) (89-91), general movements assessment (n=6) (93, 95, 98-101), standardized neuromotor examination (n=1) (106), clinical factors (n=3) (17, 18, 110) or combination of these indicators (n=3) (16, 104, 111).

Table 3.1 summarizes the characteristics of the fifteen studies that provided the diagnostic properties and/or the predictive performance of CP in preterm infants. Two additional studies were excluded; one case control study where the authors did not report the outcomes of the preterm subgroup separately (111) and another small study (30 preterm infants) that used machine learning for video analysis of general movements rather than prediction of CP (100).

The majority of the selected studies were small prospective cohorts, with duration for follow up ranging between 24-48 months, however some studies had loss to follow up rate of up to 50% (98, 99). Only two studies reported AUC (17, 93), five studies reported the classification accuracy (16, 85, 99, 101, 106), whereas the remaining studies reported the diagnostic properties, mainly sensitivity and specificity.

Thirteen of the fifteen studies traditionally used logistic regression, only two small studies used different machine learning methods, including RF, for prediction of CP from multidimensional datasets (i.e. analysis of optic flow cytometry or quantitative analysis of ultrasound images) (85, 101).

Table 3.1: Characteristics of Studies Predicting CP in Preterm Children

Study	Population CP n/N (%)	Predictors	CP Outcome	Sens	Spec	PPV	NPV	AUC	Accuracy
Logistic Regression									
Constantinou 2007 (16)	<32 weeks, <1500g CP: 10/102 (10%)	term MRI, behavioral assessment	Palisonao 24 months	80	81	36	97	-	80
Broitman 2007 (17)	< 1000 g CP: 347/2103 (16%)	All clinical model with late cUS at 36 weeks vs early clinical model with cUS at 28 days	Abnormal tone, posture or movement (Amiel-Tison) (101)	-	-	-	-	0.78 vs 0.72	-
Lacey 2004 (88)	< 30 weeks, CP: 36/203 (18%)	discharge exam vs cUS at 7 & 28 days	Abnormal motor exam 36 months	86 48	83 87	57 88	96 43	-	-
Spittle 2015 (106)	<30 weeks CP: 6/97 (6%)	combined motor tests at 4, 8, 12 months	Palisono 48 months (81)	83	95	56	99	-	92
de Vries, 2004 (86)	<33 weeks, CP: 76/429 (17%)	secquential cUS	Hagberg 24 months (85)	76	95	48	99	-	-
Woodward, 2006 (89)	<31 weeks, CP: 17/167 (10%)	MRI at term any vs mod-severe abnormality	Palisano 24 months	94 65	31 84	-	-	-	-
Nanba 2007 (90)	<34 weeks, <1500g CP: 38/289 (13%)	MRI at term for PVL vs corona radiata lesions	Palisonao 31 months	62 78	87 96	-	-	-	-
Mirmiran 2004 (91)	<30 weeks or <1250g CP: 7/61(11%)	MRI at term vs sequential cUS	Palisonao 31 months	86 43	89 82	60 33	97 87	-	-

Study	Population CP n/N (%)	Predictors	CP Outcome	Sens	Spec	PPV	NPV	AUC	Accuracy
Skiold 2013 (104)	<27 weeks CP: 4/53 (7.5%)	term MRI vs GMA at 3 months	Palisonao 30 months	100 50	98 92	80 33	100 96	-	-
Ferrari 2002 (93)	<37 weeks with abnormal cUS CP: 41/84 (49%)	Synchroniz--ed cramped GMA and Prechtl neurologic exam over first 5 ms	Palisonao 24-36 mths	79 89	100 52	100 67	84 84	0.97	-
Romeo 2008 (95)	<37 weeks CP: 57/903 (6%)	GMA score >57 vs HINE neurologic exam at 3 ms	Hagberg 24 months	98 96	94 87	-	-	-	-
Oberg 2015 (98)	<33 weeks or <1500g CP: 10/87 (12%)	GMA at 3 months	Palisonao 24 months	90	90	53	100	-	-
De Bock 2017 (99)	<33 weeks CP: 7/122 (6%)	GMA at 1&3 months	Palisonao 24 months	86	77	19	99		77
Machine Learning									
Hope 2008 (85)	<31 weeks or <1500 g (37cases, 48 controls)	cUS texture first week	Palisano 4 month	-	-	-	-		72
Stahl 2012 (101)	82 infants, gestational age not specified	optic flow cytometry at 10-18 weeks	Not specified 5 years	95	85				94

Abbreviations: AUC (area under the curve), CP (cerebral palsy), cUS (cranial ultrasound), GMA (general movement assessment), HINE (Hammersmith infant neurologic exam), MRI (magnetic resonance imaging), NPV (negative predictive value), PPV (positive predictive value), PVL (periventricular leukomalacia), Sens (sensitivity), Spec (specificity).

3.3 Prediction Modeling

3.3.1 Diagnostic Tests

A gold standard test is the ideal test(s) to diagnose a particular disease against which all other diagnostic tests are compared (112). There is no gold standard test for diagnosing CP; its diagnosis relies entirely on the clinical assessment of muscle tone, posture and movement. Validity refers to the accuracy of a test or its ultimate ability to correctly identify individuals who have a particular disease from those who do not (112). The sensitivity and specificity of a test, relative to the gold standard, is the best measure of its clinical validity. Table 3.2 highlights the various parameters that are factored in the measurement of sensitivity and specificity. To apply the table for the current study; "Disease^{+}" indicates children who actually have CP, while "Disease^{-}" indicates children who do not have CP, whereas, "Test^{+}" and "Test^{-}" refers to the predicted model (or test) of those children with and without CP, respectively.

Table 3.2: Diagnostic Properties of a Test

True Class				
Predicted Class		Disease^{+}	Disease^{-}	Total
	Test^{+}	a True positive (TP)	b False positive (FP)	a+b Total test^{+}
	Test^{-}	c False negative (FN)	d True negative (TN)	c+d Total test^{-}
	Total	a+c Total disease^{+}	b+d Total disease^{-}	a+b+c+d Total population

1. True positive (TP): individual with a positive test and has the disease

2. False positive (FP): individual with a positive test but does not have the disease

3. True negative (TN): individual with a negative test and does not have the disease

4. False negative (FN): individual with a negative test but has the disease

Sensitivity measures the ability of the test to correctly identify those with the disease Therefore, a highly sensitive test if negative is useful for ruling out the disease

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{true\ positive}{true\ positive + false\ negative}$$

Specificity is the ability of the test to correctly identify those free from the disease Therefore, a highly specific test if positive is useful for ruling in the disease

$$\text{Specificity} = \frac{d}{b+d} = \frac{true\ negative}{true\ negative + false\ positive}$$

Sensitivity and specificity are inversely proportional, and they are independent of the population tested (112, 113). Serious but treatable diseases require a test that is highly sensitive (e.g., cancer screening test). However, specificity would be compromised resulting in unnecessary anxiety and unwarranted further investigations (113). The predictive values are useful in clinical medicine when considering the value of a test for a clinician, because they answer the question of how likely an individual with a positive test to have or develop the disease or the outcome of interest (113).

The positive predictive value (PPV) is the probability of having the disease when the test is positive.

$$\text{PPV} = \frac{a}{a+b} = \frac{true\ positive}{true\ positive + false\ positive}$$

The negative predictive value (NPV) is the probability of not having the disease when the test is negative.

$$\text{NPV} = \frac{d}{c+d} = \frac{true\ negative}{true\ negative + false\ negative}$$

Unlike sensitivity and specificity, PPV and NPV are closely related to the prevalence of the disease in the population. Assuming that all other factors remain constant, as the prevalence declines, the PPV decreases while the NPV increases (112, 113).

Accuracy is the ability of the test to correctly identify those with and without a disease or outcome of interest.

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{true\ positive + true\ negative}{true\ positive + false\ positive + true\ negative + false\ negative}$$

Likelihood ratio (LR) is used to determine the usefulness of a test by comparing its sensitivity and specificity or the ratio of its true positive rate to its false positive rate. The positive LR is how likely the test result being positive in an individual with a specific disease or outcome of interest compared to the same test result being positive in an individual without the disease or outcome; if the test result would change the probability of having or developing a disease in an individual. Therefore, the higher the positive LR the better the classifier, whereas the lower the negative LR, the better the classifier (113).

$$LR^{+} = \frac{TP\ rate}{FP\ rate} = \frac{Sensitivity}{1 - Specificity}$$

$$LR^{-} = \frac{FN\ rate}{TN\ rate} = \frac{1 - Sensitivity}{Specificity}$$

Receiver operating characteristic curves (ROC)

The ROC curve is used to assess the ability of a diagnostic or prognostic test to identify individuals who have or will develop a given disease. It is a plot of the true positive rate of a test (sensitivity) on the y-axis against its false positive rate (1- specificity) on the x-axis, for every possible cut-off point (113, 114), hence depicting the trade-off between the sensitivity and the specificity of the test of interest. For logistic regression models, the ROC curve evaluates the discriminative or classification performance of the model; the ability of the model to identify individuals with or developing a given disease at all possible cut-off points. Figure 3.1 shows that the closer the ROC curve to the ideal point (top left corner with 100% sensitivity and specificity), the better the performance of the classifier (113).

The area under the curve (AUC) represents the overall accuracy of a test to discriminate between those who have or will develop a given disease from those who do not, over all possible thresholds or cut-offs. The AUC ranges from 0.5 to 1. The larger the area the better is the classifier; where 1 is the optimal test or classifier (100% specificity and 100% sensitivity) and 0.5 is a worthless test or classifier (50% specificity and 50% sensitivity) that is not different from flipping a coin. In Figure 3.1, the upper curve (C) represents a test with high sensitivity and specificity and AUC approaching 1.0, while the dotted A line represents the line of discrimination with an AUC of 0.5. The AUC of most tests used in health research lies between these two extremes (curve B).

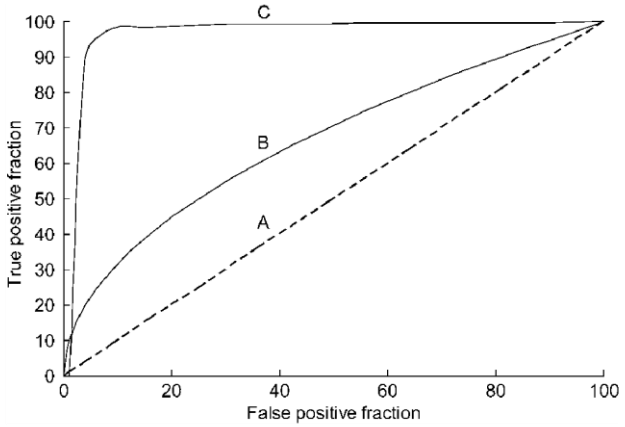

Figure 3.1. Receiver Operating Characteristics Curve

With permission from: Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity.[113] (A) line of 0 discrimination (AUC = 0.5); (B) typical clinical test (AUC = 0.5-1.0); (C) perfect test (AUC = 1.0)

For prediction models, ROC curves can be used for identification of an optimal classifier because they provide the cut-off point at which sensitivity and specificity are maximized relative to one another. The ROC curve is also useful to compare different classifiers (models) based on the AUC that provides an unbiased measurement of the performance of prediction across different models (114).

3.3.2 Logistic Regression

One of the main goals of epidemiological research is to examine the association between an exposure and an outcome of interest. Measures of association can be assessed using absolute measures (attributable risk or risk difference) or relative measures (RR and OR). Logistic regression has been traditionally used in health research to assess the relation between independent observations/exposures and a binary outcome both for association and for prediction.

Risk prediction models are used to estimate the risk of having a specific prevalent disease (diagnostic) or developing one (prognostic). Multivariable prediction models are one method to estimate the probability of a certain outcome in individuals, given their set of predictors (115, 116).These models can be developed from prospective cohort, randomized clinical trials, or nested case-control studies using both categorical and continuous explanatory variables (116). For a binary outcome variable, such as CP, logistic regression provides the odds of CP (the ratio of the probability of developing CP to the probability of not developing CP) using the following formula:

Odds of CP = P/(1-P) Probability of CP (P) = $e^{(a+bx)} / 1+ e^{(a+bx)}$

In logistic regression, the OR or the ratio of the odds, represents the constant effect of a predictor X on the likelihood that outcome Y (CP in this study) will occur (115). Whereas with probability, the effect of a predictor X on the probability of the outcome Y is not constant and has different values depending on the value of X.

Risk prediction models are increasingly used in clinical medicine as adjuncts to guide clinical reasoning and decision-making, provided that accurate estimates of the probability risk and validation, both internal and external, were performed. Moons et al proposed important steps for development, validation and reporting of risk prediction models. The following section on developing a multivariable model and assessing its usefulness and its performance is derived from the recommendation of the first series of Moons et al (116). External validation of the developed model is based on the second series of Moons et al (117).

Model Development

Prior to developing the logistic regression model, both the outcome of interest and the predictors should be clearly defined and measured in a standardized and reproducible way. The selection of relevant predictors is usually based on expert opinion derived from scientific knowledge of the outcome of interest. Careful selection of predictors is critical and should include not only causal but all potential correlates. However, attention should be paid to the number of predictors to be included in the model, using the rule of 1 to 10 event per variable; at least 10 individuals having (developed) the outcome of interest are required per 1 predictor in the model (118). For the outcome, the methods used to measure and ascertain the outcome, independent of or blinded to the studied predictors, and the duration of follow-up should be clearly defined.

Analysis of Logistic Regression Model

- Missing values: Multiple imputation of missing data may be performed to avoid bias if the analysis included only individuals with completely observed data.
- Continuous predictors: Testing for linearity and using simple transformation of non-linear continuous predictors increases the predictive ability of the model.
- Predictor selection in the multivariable model is preferably performed using a full model approach or backward selection procedure, rather than inclusion of only univariate correlates.
- Logistic regression provides regression coefficients (Wald test or LR test) that assess the relative contribution or the relative weight of each predictor in the model (the effect of one unit change in the predictor X on the outcome Y, when all other predictors were kept constant). The model's intercept quantifies the baseline risk for an individual to develop the outcome of interest, if all predictor values were zero (115, 116).
- The linear predictor from the logistic regression model can be transformed to provide a predicted probability. The ROC curve is used to determine the optimal cutoff. The model performance can be assessed quantitatively using discrimination (AUC), calibration (plots), and classification (115, 116).

Internal Validation

The probability of developing an outcome as predicted by the model compared to the observed one, provides the 'goodness fit' of this prediction model. By design, prediction models are expected to optimally fit the development sample. However, they are less accurate when tested in new, but similar, individuals due to overfitting to the development sample. The potential for overfitting is indirectly proportional to the number of outcomes/events in the development sample and is directly proportional to the number of predictors (relative to the number of events) (116). To test for overfitting of prediction models, internal validation can be performed using different methods:

1. Split sample: The training data is randomly split once into two subsamples: one to develop the model and the other to test its performance. Splitting can be done as split half or other fractions. The measures of performance are based on similar but independent data from the same population. However, this method is inefficient because of data wasting (119).
2. Cross-validation: This is an extension of the split method where the model is developed on several random splits of the data; within each split, one part serves as the training set and the other part serves as the validation or testing set. The performance of the model at each split is calculated and the process is repeated, with one data subset left out at a time, until all subsets served once to test the model (e.g. if within each random split, 90% of the data are used to develop the model and 10% to test it, then the process is repeated at least 10 times). The average performance of the final model is then calculated from all the splits and the stability of the cross-validation improves with more repetitions (100 times) (119).

3. Bootstrapping: This method uses an intensive computer-based resampling technique, or drawing with replacement, of around 100-500 subsamples, from the original sample and each bootstrap sample is of the same size as the original sample. The prediction model may be developed from the bootstrap sample and tested on the original sample, or vice versa, so that 100% of the data is used for model development and 100% for model testing. The estimates of performance can be assessed on each bootstrap subsample and the final model accuracy can be then computed by the average performance of the prediction models. Compared to the split methods, bootstrapping is considered to be a more preferred method for validation as all the data set will be used without wasting. It is considered to be the most efficient validation method, particularly when the sample size is small or with a large number of predictors (119, 120).

However, the main disadvantage of the cross validation and the bootstrapping approaches is their computer-based automated model selection procedure that does not allow exploration of the data or the use of judgment during the selection of predictors.

External Validation

External validation is a method used to assess the predictive performance of a previously developed model when applied to a sample that is temporally or geographically different from the development sample and it is considered as a marker of "generalizability" of the developed model when applied to different populations (117). The process includes: taking the original model with its predictors and regression coefficients, assessing the predictor and outcome measures in the new population, applying the original model to these new data and finally assessing the model's predictive performance quantitatively as outlined before (117). As expected, external validation often results in lower performance than internal validation, when the developed model is applied to different individuals or different populations.

Limitations of Logistic Regression for Developing Prediction Models

Traditionally, logistic regression has been used for prediction of outcomes in health research, but it has many drawbacks. Logistic regression model requires correct specification of the main effects and the interactions between predictors, otherwise the resulting prediction may be biased. Additionally, the model assumes a linear relationship between the predictors and the outcome (116); ignoring a non-linear relationship would result in a poorly fitted model. To overcome these limitations, machine learning methods have been increasingly used for prediction and probability estimation for genomic, genetics, biomedical and medical research (121-127), and in clinical epidemiology (128). They do not require specification of the underlying model and can handle complex and non-linear relationships between the predictors and the outcome (22, 23). They are often used in settings with a large number of predictors relative to the number of observations, such as genomic research, which would pose a problem if logistic regressions were used. Machine learning methods are used to develop models that predict an outcome from a set of predictors and can rank the predictor variables based on their relative importance (weight) for prediction (22, 23). In the following section, I will discuss certain types of machine learning methods, namely, decision trees and RF ensemble methods. Building prediction models from decision trees and RF will be described and the difference between RF and logistic regression will be highlighted (see also Table 3.3).

3.3.3 Decision Trees and Ensemble Methods

Decision Trees

The Classification and Regression Trees (CART) method was first introduced by Breiman in 1984 and provided solutions for regression and classification problems that are easily interpreted with a visual graphic display as inverted trees. The CART, also known as binary recursive partitioning, are broadly grouped into CT for binary or categorical outcomes and regression trees for continuous outcomes (22).Distinguishing “CP” from “non-CP” can be regarded as a classification problem within this learning field.

The CART splits a whole sample in hierarchical manner, starting from the root node and growing branches using a sequence of binary split rules for the explanatory variables. Each binary split maximizes the discrimination of the outcome between the resulting child nodes and maximizes the homogeneity between participants within each child node that is known as node purity (127). This split criterion minimizes the residual sum of squares (for regression trees) or the Gini index (for CT)) (22). The branching or splitting continues until a stopping point is reached or no more splits are possible; these are called the terminal nodes. Each terminal node is assigned a predicted numerical value (regression) or a predicted outcome category (classification). Once all terminal nodes are identified, the CART can be used to provide a prediction for regression (the mean response in each terminal node of the tree) or classification (the majority vote or the most commonly occurring class in each terminal node of the tree). The tree is then subject to pruning, a process by which cross-validation generates nested trees from a training data set and selects the optimal final tree when applied to a testing data set (22). As this study investigates CP as a binary outcome, the remaining section will focus on CT and RF building, performance, and diagnostic properties.

Classification Tree

Building a prediction model using CT follows the same previously described steps for building CART with the final decision tree being an inverted tree composed of yes and no answers at each split (126). A decision tree can be easily interpreted by healthcare providers and its output is similar to the clinical reasoning process or clinical algorithms commonly used to guide patient management. The advantage of a CT in healthcare is its ability to clearly identify subgroups of patients who are at the highest risk of developing the outcome of interest, represented as the terminal nodes of CT.

The performance of a CT is dependent on the number of explanatory variables, the size of the tree grown (number of splits, leaf size), and the split criterion (22, 126). The performance of CT can be assessed by calculating how correctly CT is able to classify those with the outcome of interest (sensitivity) and those without (specificity). Additionally, penalties for misclassification may be used to improve the accuracy of the prediction (22).

While decision trees are fairly easy to implement, understand, and interpret, they have limitations. First, the splitting algorithm is "greedy", so that the built tree is optimal at each split but may not be optimal globally. Secondly, the tree is “unstable” where slight changes in the data may result in a substantially different tree. Thirdly, the algorithm tends to over-fit to the training data, resulting in a much weaker performance in a testing data set. The predictive accuracy and robustness of decision trees can be improved by aggregating many decision trees using the RF ensemble method (23).

Random Forest

To overcome the limitations of prediction using the decision trees, Breiman (2001) proposed the RF algorithm, by growing a forest or ensemble of CT and letting them vote for the most popular classification. In comparison to a single tree, using randomness in building each tree in the forest leads to a better prediction and does not have the problem of overfitting (23). The RF develops a large number of trees (hence the name forest) in bootstrapped samples of the whole data set. When building each tree within the forest, only a random subset of predictors is available at each node to create a split (the random part of the forest) (23). The tree-growing process is repeated until a preset number of trees are reached. The ensemble of trees then "votes" on the optimal classification or the majority vote (mode of the classification of the individual trees). The predicted outcome is estimated by the most frequent predicted outcome from each component tree (for classification) (23). Important tuning parameters of the RF predicted model include: the number of trees, the size of the terminal node, the number of features available at each split and the number of predictors used at each node in growing each tree (23). The diversity of the trees helps to improve the accuracy and stability of the prediction, because the aggregate vote of several decision trees is less susceptible to noise and outliers than a single tree (23). A main advantage of RF is providing information on what variables are important in the classification by computing an importance score for each variable, based on how much their presence in the forest improves the prediction compared to a model without the variable. This score can be then used to rank variables relative to each other (23).

A fundamental concept in RF is the out-of-bag (OOB) sample, which refers to the set of observations that were not included in the bootstrap sample, corresponding to approximately one-third of the original data set (23). Each tree within a RF uses a different bootstrap sample and therefore a different OOB sample. The OOB sample is then used to evaluate each component tree within the forest by estimating the generalization error, defined as the error rate of the OOB classifier on the training set (23).

Classification using RF is a useful tool for prediction of a binary outcome because of its discriminative ability. However, estimating the probability of an outcome (risk) is important in clinical medicine. Recently, RF has been used as a probability machine to estimate the conditional probability for binary outcomes and providing risk estimates and effect size (OR with their 95% confidence intervals) as well as the interaction effects between predictors (124, 129, 130).

3.3.4 Comparison of Random Forest and Logistic Regression

Both RF and logistic regression have been used for risk prediction in health research. However, they differ substantially in the method of model development and the output. Logistic regression is explanatory; it provides regression coefficients that determine the relative contribution of each predictor in the model, when other predictors are kept constant. Whereas, RF is like a "black box" that only provides the variables selected in the model and their importance ranking. Table 3.3 highlights some of the main differences between risk prediction using multiple logistic regression and RF ensemble methods.

Table 3.3: Comparison of Multiple Logistic Regression and Random Forest.

Methodological issues	Multiple logistic regression	Random forest
Sample size	A large sample size is required to provide a sufficient number in both categories of the outcome variable.	No minimum sample size
Effect estimates	OR and their confidence intervals.	By default, only variable importance rankings are produced. Risk ratios and their confidence intervals may be produced using additional steps (128).
Selection of explanatory variables	The number of explanatory variables must be selected first and they should not exceed 10% of the events number.	All available variables may be used. When building each tree, only a random subset of predictors is used.
Linearity assumption	Assumption of linear relationship between each explanatory variable and the logit of the outcome variable.	No assumption about the shape of the relationship between the explanatory and the outcome variables.
Distributional assumptions	No assumptions about the distribution of the explanatory variables.	No assumptions about the distribution of the explanatory variables.
Dealing with complex interactions	Interactions (especially higher order ones) between explanatory variables are difficult to identify and interpret.	Can deal with higher order interactions but does not explicitly identify them in the final model.

Adapted from Henrard et al.(126)

Recent studies compared conventional regression models (logistic or Cox-regression) to different machine learning methods (RF, neuronal network, support vector machines and gradient boosted decision trees) for prediction of clinical outcomes. These studies reported superior prediction (AUC) of machine learning methods over conventional regression or clinically-based risk scores in the adult population (131-135). Examples of clinical outcomes predicted by these studies included: cardiac complications in patients with acute chest pain (131), re-hospitalization in patients with heart failure (132) , mortality in patients with sepsis (133), disposition of adults with acute obstructive airway disease (134), and triaging of adult patients in Emergency Department (135).

There is emerging literature comparing machine learning methods to conventional logistic regression in perinatal, neonatal or even pediatric research. Recently, Goto et al (2019) reported superior performance of RF over conventional regression in prediction of critical care and hospitalization among children admitted to the emergency departments (136). Similarly, Carlos Campillo-Artero et al (2018) studied data of 6,157 singleton births and reported superior performance of RF over logistic regression for predicting emergency Cesarean section (AUC 0.94 (95% CI: 0.93–0.95) vs AUC 0.78 (95% CI 0.76–0.8)) (137). However, other studies reported inferior performance of RF compared to logistic regression when clinical predictors were used with lower AUC and worse predictive accuracy. Kuhle et al (2018) conducted a population-based study of 30,705 singleton infants comparing logistic regression to machine learning methods for prediction of fetal growth abnormalities. The study showed that machine learning methods did not add advantage to the conventional logistic regression and reported poor prediction (AUC 0.6–0.7) for primiparous women and fair prediction (AUC 0.7–0.8) for multiparous women, irrespective of the method used (138).

In summary, recent studies comparing machine learning methods, including RF, to conventional logistic regression for their discriminative ability in prediction of clinical outcomes reported conflicting results from superior to similar or even inferior performance. The conflicting results suggest that none of the prediction methods is superior and the predictive accuracy may rely on the settings or the datasets and may not be constant across different studies. The main advantage of machine learning over conventional regression is its ability to handle large/multidimensional datasets (where the computer can learn iteratively from the data to develop prediction) or complex datasets (with non-linear relations between predictors and outcome or interactions between predictors). However, their predictive performance compared to conventional regression outside these settings, is yet to be explored.

3.4. Relevance and Rationale

Cerebral palsy is the most common physical disability in children worldwide. Although the perinatal insult of CP is considered static, the associated disability is progressive, and its severity increases over time. In addition to spasticity, weakness and immobility, CP is commonly associated with major comorbidities (epilepsy, behavioral disorders, developmental delay, dysphagia or tube feeding, hip dislocation and muscle contractures). This disabling disease has a lifelong impact on the general health of affected children, their societal integration and quality-of-life. Additionally, it burdens their families, society and the healthcare system with both healthcare cost and services utilization.

The risk of CP in preterm children is 10 times higher than that of the general population, with extreme preterm children being at the highest risk (8). As early diagnosis and intervention are important in the management of CP, timely identification of preterm children at risk of CP is desirable to benefit from rehabilitation programs with the potential to improve their functional outcomes and quality of life (11, 27, 74-76).

There is a need for reliable tools to accurately predict CP in preterm children early on, as the approach for diagnosis and management is challenging; (i) there is no specific test to diagnose CP and the diagnosis is solely dependent on detailed neurologic examination which may delay the diagnosis if not performed timely and by expert clinicians, (ii) the criteria for referral of high risk preterm children to intervention programs varies across countries including Canada; depending on the jurisdictional regulations and healthcare resources, (iii) the time to diagnose CP is variable and depends on the access to specialized care, particularly outside hospital settings or follow up programs, where the clinical diagnosis may be delayed beyond the commonly established time around 18 to 24 months of age, (iv) by 5years of age, almost one third of children with CP would have already established comorbidities which impact their future health outcomes and quality of life (26), (v) the clinical outcomes and healthcare cost have a direct relation with the disease severity and the associated comorbidities, both have potential to be ameliorated if rehabilitation and interventions were started early within the first 1-2 years of age, at the time of brain plasticity, with the potential for neuronal recovery and improve mobility and motor functions (11, 27, 74-76).

Research that focuses on developing accurate prediction models of CP in this high-risk population is very valuable for patients, their families and healthcare providers. Accurate prediction of CP in preterm survivors is crucial to enable early identification and to guide individualized interventions with the potential for neuronal recovery and improvement of mobility and quality of life. Multiple studies showed that the severity of CP and the associated comorbidities are modifiable, if the children at risk are identified early and referred to the appropriate rehabilitation services. This underlines the importance of early implementation of accurate tools to identify those preterm children at the highest risk of CP. In addition to improved mobility, improvement of other developmental domains such as language, cognitive and problem-solving skills have been reported with early intervention in preterm children with CP (11, 27, 74-76). Families of those children engaged in such programs were shown to have enhanced parenting skills, improved bonding and parent-child interaction and above all improved children's engagement in social activities. Such modifications of the neurosensory stimulation and the environment around CP children have an impact on their behavior, communication and social integration (77, 78).

For families, accurate prediction is invaluable when counselled about the risk of their children developing CP, or NDI, after preterm birth or a significant perinatal event. Interaction between parents of preterm children and healthcare providers constitutes an integral part of their experience in NICUs and during antenatal counselling prior to preterm birth; such a stressful situation reportedly impacts families' coping with the trauma of preterm birth, particularly related to parent child interaction and parental empowerment.

For caregivers, early and accurate prediction of CP may assist in selecting the appropriate treatment and providing individualized interventions for this high-risk patient population. It would also enable targeting the necessary resources to those children at the highest risk who would most likely benefit from early detection and intervention.

Finally, logistic regression has been traditionally used for risk prediction in neonatal literature. The emergence of novel prediction methods, that overcome the limitations of logistic regression, has stimulated researchers to explore and compare the predictive accuracy of these methods to conventional logistic regression. The method with the highest accuracy and ability to discriminate between those children at risk for an outcome of interest from those who are not can be implemented in clinical practice.

The predicted probability of an outcome of interest, such as CP, for each preterm infant given their set of predictors, can be easily transformed into calculated risk based scores and algorithms. Clinical calculators of these algorithms or risk-based scores derived from such prediction models have been widely used in neonatal practice to aid counselling or when critical decisions are discussed. Examples of such clinically available calculators include:

(i) NICHD calculator for prediction of death or disability when counselling families regarding active resuscitation at the edge of viability or to provide prognostication of long term outcomes after extreme preterm birth, https://www.nichd.nih.gov/research/supported/EPBO/use (139)

(ii) The BPD calculator to decide for selective administration of systemic steroids among very preterm infants with severe RDS, a therapy known to increase the risk of CP so selectively given to the sickest infants, https://neonatal.rti.org/index.cfm (140).

(iii) Vincer et al developed an algorithm for prediction of mortality at extreme preterm birth, using the same population database of the current study (141).

Knowledge Gap and Rationale

Despite considerable evidence in the literature on the risk factors associated with CP in preterm children, few studies have advanced beyond risk analyses to develop risk prediction models and measure their diagnostic properties. Moreover, the majority of these reports were limited by small sample sizes and not being population-based (Appendix 1). Additionally, multiple studies did not investigate CP as the main outcome, but rather examined NDI, with CP as one component of this composite outcome. Traditionally, these studies used logistic regression for predicting CP in preterm children.

The role of machine learning methods, such as the RF, in the prediction of CP in the preterm population has not yet been adequately explored. The RF method is one of the most commonly used and accurate machine learning methods. It does not require a model's specification and can handle complex relationships between predictors and the outcome. Therefore, RF may result in better prediction of CP when compared to logistic regression. Additionally, it is not known if the claimed superiority of RF over logistic regression would remain when clinical predictors are used. Therefore, the proposal of this study is to explore the role of RF in prediction of CP using clinical predictors and to compare RF to conventional logistic regression for the accuracy in prediction of CP among a large population-based cohort of very preterm children in Nova Scotia. To my knowledge, no study compared RF to logistic regression in this context to date.

As early identification and referral of preterm born children at risk of CP has been shown to improve their outcomes and quality of life, research that focuses on reliable risk prediction in this population is needed. With advances in perinatal care and improved survival of very preterm children over the last few decades, we expect to see more children at risk of CP which makes a reliable prediction tool more pressing now than ever before.

CHAPTER 4: METHODS

4.1. Study Population and Design

This study used retrospective data of a prospective population-based cohort of very preterm infants. The birth cohort includes all live born very preterm infants ($22^{0/6}$ - $30^{6/7}$ weeks gestation) who were born between January 1, 2000 and December 31, 2014, to mothers residing in Nova Scotia. From this birth cohort, only infants with no major congenital anomalies or no palliation at birth were eligible for inclusion. Children who died before 36 months of corrected age, those who were lost to follow-up, those with missing outcome data and those who had their last assessment before 18 months of corrected age were excluded. Data of surviving very preterm infants who received standardized neurodevelopmental assessment up to 36 months corrected age were collected from the Nova Scotia PFUP database and were evaluated for the primary outcome of CP.

Candidate variables were classified into three groups according to pre-specified time-points in a chronological fashion: prenatal period (maternal, pregnancy and fetal factors), perinatal period (factors related to intrapartum period up to 6 hours after birth), and postnatal period (short term morbidities up to hospital discharge at or near the expected date of delivery). For the purpose of this study, we defined those three time points being relevant for prediction of CP both for families and caregivers. Accurate prediction of CP at these three phases are crucial to aid in counselling families for prognostication or when informed critical decisions are discussed: (i) before birth (to decide if active resuscitation will be provided for extremely preterm infants), (ii) within the first 6 hours after birth (to decide regarding continuation or withdrawal of intensive care if a catastrophic perinatal event occurred), and (iii) at hospital discharge (to assess the long-term prognosis and to guide referral of high-risk infants to early intervention programs and other developmental services).

Before analysis, the full dataset was randomly divided into training and testing subsets in a 70:30 ratio. Prediction models of CP using logistic regression and RF were developed in

the training dataset and were validated in the testing dataset. The internal validation was performed to assess for model overfitting and to examine how the model would perform in a similar but independent sample from the same population. For internal validation, the splitting method was preferred over K-fold cross-validation or bootstrapping as the model building procedure for logistic regression requires human intervention that cannot be easily automated. The validated prediction models were compared with regard to their discriminative ability (AUC), accuracy (correct classification) and diagnostic properties (sensitivity, specificity, PPV, NPV, LR).

4.2. The AC Allen Provincial Perinatal Follow Up Program Database

The AC Allen Provincial PFUP database is the data source for this study. Since 1993, the PFUP has enrolled all liveborn very preterm (<31 weeks) infants born to mothers who resided in Nova Scotia. The database collects a broad range of data including sociodemographics, prenatal, perinatal, and postnatal clinical data as well as neurodevelopmental data up to 36 months of corrected age. The database also collects data on delivery room deaths at any hospital in Nova Scotia. The PFUP database contained records on all 1111 very preterm infants born to mothers residing in Nova Scotia between 2000 and 2014. The Program Medical Director (Dr. Michael Vincer) and the database manager perform periodic audits and code checks of the database including retrospective updating of coding schemes (in case of changes in coding definitions) in order to maintain the database consistency.

The PFUP performs a standardized neurodevelopmental assessment of all surviving preterm children up to 36 months of corrected age with a follow-up rate around 96%. Following discharge from NICU, all very preterm infants are scheduled for visits to the PFUP clinic at 4, 8, 18, and 36 months of corrected age. If an abnormality is detected at any visit, more frequent follow-up may be required. Each infant is assessed during those visits by a multidisciplinary team including pediatricians, nurses, physiotherapist, occupational therapist, dietician and speech therapist. Each visit includes a complete history and physical assessment including a detailed neurologic examination. If serial neuromotor

assessments are completely normal up to 18-24 months of corrected age, no further neurologic exam is required, but detailed developmental assessment of motor skills continues up to 36 months of corrected age using standardized validated tests as described below. However, if an exam shows an abnormal finding or parental concerns were identified at the 18-24 months visit, then a detailed neurologic exam is conducted at each subsequent visit up to the 36 months to assess for milder forms of CP.

Additionally, standardized neurodevelopmental screening tests are conducted at each PFUP visit to assess gross motor functions. The Alberta Infant Motor Scale (AIMS) is performed up to 8 months of corrected age (107) and the adapted Bayley Scale of Infant and Toddler Development (BSITD) screener is used to evaluate both fine and gross motor skills beyond 8 months of corrected age (142). The cognitive and language development are also assessed at each visit using the adapted CAT/CLAM (Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale) (97, 143). These validated tests were shown to predict motor and cognitive/language outcomes, respectively (97, 107, 142, 143). Finally, standardized developmental testing (BSITD, Edition II-III), conducted at 36 months of corrected age, provides psychomotor and mental developmental scores (cognitive and language development) that are widely used in longitudinal studies of preterm infants (144, 145).

If CP is suspected, the child is referred to a pediatric neurologist to confirm the diagnosis and to initiate the management, including referral to a rehabilitation program. Once the diagnosis is confirmed, children with CP are further classified based on ambulation into mild CP (level 1-2) or moderate to severe CP (level 3-5) using the Palisano GMFCS (83). The GMFCS is a validated tool that has been used extensively for classification of CP in literature and to standardize reporting of CP allowing comparison of these studies in meta-analyses and systematic reviews (Appendix 1).

4.3 Outcome

The outcome of the study is the presence of CP of any severity, defined as a disorder of control of movement, muscle tone and/or reflexes or posture secondary to a non-progressive brain lesion. The prevalence of CP was determined at 36 months of corrected age, rather than at 12 or 24 months, as the diagnostic accuracy is higher, particularly for the mild form or the ambulatory subtype (24, 25). If the assessment at 36 months is not available, then the most recent standardized assessment close to 36 months (range 18-42 months) was used to determine CP.

The clinical diagnosis of CP at the PFUP is described above. Cerebral palsy is coded in the PFUP database in three ways: as a binary categorical variable (yes/no), as a nominal categorical variable based on the clinical subtypes (spastic, athetotic, hypotonic, or ataxic), and as an ordinal variable based on the standardized GMFCS level. The definition of CP, the primary outcome of this study, has been standardized in the PFUP database since its inception and the severity classification of GMFCS levels has been standardized in the PFUP database since 2000.

4.4 Candidate Predictors

Potential predictors of CP in this analysis included the maternal and infant factors described in Section 3.1.4, or other variables that are associated with CP (e.g., maternal age) or have biologic plausibility or are antecedents to CP (e.g., birth asphyxia, infection, or brain injury). The complete list of the predictors with detailed information of their corresponding codes and definitions is provided in Appendix 2.

Prenatal factors included (i) maternal factors (age, parity, socioeconomic status as per Hollingshead classification, single parent, previous neonatal deaths or previous stillbirths); (ii) pregnancy factors (exposure to smoking, alcohol or drugs during the current pregnancy, gestational hypertension/pre-eclampsia, diabetes, idiopathic preterm labor, tocolytics, pre-labor premature rupture of membranes, chorioamnionitis, antepartum hemorrhage); and

(iii) fetal factors (multiple gestation, fetal growth restriction, fetal distress).

Perinatal factors included (i) intrapartum factors (intrapartum magnesium sulfate, antenatal steroids, mode of delivery, gestational age, birth weight, SGA, infant sex, outborn status, birth asphyxia, need for resuscitation (chest compression or epinephrine) at birth, and 1- and 5-minute Apgar scores); and (ii) factors related to the first 6 postnatal hours (infant's body temperature, hemoglobin, and blood pressure on admission to the NICU). Birth weight z-scores were determined relative to a Canadian reference population to assess infant growth independent of gestational age and sex (146).

Postnatal factors included (i) postnatal treatments (surfactant, prophylactic indomethacin, ibuprofen or indomethacin for medical closure of patent ductus arteriosus, inhaled nitric oxide, antireflux medications, muscle relaxants and systemic dexamethasone); (ii) neonatal morbidities related to preterm birth (hypoglycemia, hyperglycemia, anemia, thrombocytopenia, respiratory distress syndrome, severe IVH (defined as grade 3 or 4 IVH as per Papile classification) (51), parenchymal echodensities, cerebral white matter cystic lesions (PVL or porencephaly), BPD (defined as oxygen dependency for at least 28 days with cystic changes on chest x-ray), severe ROP (defined as stage 3 or higher based on the revised international classification system (147) or requiring intervention), NEC (defined as Bell stage 2 or higher) (148), sepsis (defined as positive bacterial, viral, or fungal blood or cerebrospinal fluid culture); (iii) severe neonatal illnesses (resuscitation during NICU stay, pneumothorax, inotropes for hypotension or cardiac dysfunction, major surgery and discharge on home oxygen); and (iv) the number of days on mechanical ventilation and length of hospital stay.

4.5. Data Preparation

Data coding, checking for missing values, assessment of distribution of continuous variables, and testing for linear relationships between continuous variables and the logit of CP were performed.

Three observations with unusual clinical courses were removed from the data (1 child with extremely long hospital stay and 2 children with extremely long mechanical ventilation days). Biologically implausible values were replaced with missing values. Patient's records were checked for extreme values for recording or abstraction errors and no abstraction errors were identified. Extreme values confirmed to be true values were retained and those with presumed error on recording or abstraction were replaced by missing values.

Data coding included dichotomization of some continuous variables: number of cigarettes smoked/day to maternal smoking during pregnancy, duration of rupture of membranes to prelabor premature rupture of membranes >18 hours, and duration of nasal continuous positive airway pressure (CPAP) to nasal CPAP. Some continuous variables were used to create new categorical variables: gestational age in weeks to create extremely low gestational age (<26 weeks), birth weight in grams to create extremely low birth weight (<1000 g), duration of antenatal steroids to create appropriate antenatal steroids (>24 hours prior to delivery), and mean blood pressure to hypotension on admission to NICU (If mean blood pressure value is less than gestational age at birth). For variables with multiple codings/definitions (e.g. maternal diabetes), the most accurate definition or the one with the lowest missing data was selected.

New variables were created for any maternal hypertension (to include pre-existing hypertension, gestational hypertension or pre-eclampsia), neonatal inotropes (to include receipt of dopamine or dobutamine) and cystic white matter lesions (to include cystic PVL and porencephalic cyst). There was a change in practices over time for PDA medical treatment (from indomethacin to ibuprofen) and for ROP treatment (from laser surgery to intravitreal injection of bevacizumab); therefore, new variables were created to reflect any

treatment regardless of the year of birth: PDA medical treatment (ibuprofen or indomethacin) and any ROP treatment (laser surgery or bevacizumab).

Checking for missing data revealed 16 variables with missing values in the dataset; the list of those variables is provided in Appendix 2. No imputation for missing variables was conducted, as the proportion of missing values was low. Only two variables had equal or >5% of missingness (maternal smoking 5% and SES 11%), and a "missing" category was created for these two variables. Pairwise deletion was used for model development where observations with complete data for predictors were included in the model, whereas those with missing data for other predictors were excluded.

Histograms were created to assess the distribution of continuous variables; maternal age, birth weight, z-scores of birth weight, admission temperature and admission hemoglobin were normally distributed (Appendix 2).

Testing for linear relations between continuous variables and the logit of CP was conducted. The continuous variables, blood pressure on admission, duration of CPAP, and duration of high frequency oscillatory ventilation, were dichotomized as they showed a non-linear relation with the logit of CP. The relation of the remaining continuous variables was reasonably linear over data-dense sections of the independent variables (Appendix 2).

4.6 Statistical Analysis

Comparison of variables between the groups of children with and without CP in the full dataset was done using Fisher's exact tests for categorical variables and t-tests or Mann-Whitney tests for continuous variables as appropriate. Then, the full dataset was randomly divided into a training (70%) and a testing (30%) datasets prior to analysis. Stata/IC 16 (Stata Corp., College Station, TX, US) was used for all analyses with the exception of the RF models.

Objective 1: Logistic regression model

Model building. Using the training dataset, candidate predictors of CP were identified using a series of logistic regression models. First, unadjusted models were run for each candidate predictor of CP. Then, three multivariable regressions were built using the candidate predictors available at each of the prespecified time points (prenatal, perinatal, and postnatal). Variables with $P < 0.1$ in the unadjusted analysis were entered into the corresponding multivariable model, and variables with $P < 0.05$ were removed from the model (149).

The set of candidate predictors from these three logistic regression models developed at each time point were then added incrementally together and only variables with $p < 0.1$ were retained, resulting in three multivariable regression models as follows:

- Prenatal Model: Maternal and pregnancy-related variables.
- Prenatal/Perinatal Model: Prenatal Model plus perinatal variables (intrapartum and early neonatal variables within 6 hours from birth).
- Full (Prenatal/Perinatal/Postnatal) Model: Prenatal/Perinatal Model plus postnatal variables up to hospital discharge.

Model testing. For each of the three models that were developed, the predicted probability of developing CP for each participant given their set of predictors was computed in both the training and testing datasets using (115)

$$P\,(CP = 1) = e^{(a+b1x1+b2x2+\ldots.bixi)} / 1+ e^{(a+b1x1+b2x2+\ldots.bixi)}$$

Performance of the logistic regression models at the three time points were assessed for discrimination, calibration, and classification. The diagnostic properties of the models were determined (149).

Discrimination of the models was assessed using c-statistics or AUC to estimate the overall ability of the prediction models to discriminate between children who develop CP from those who do not. The AUC, based on the trade-off between the true positive (sensitivity) and the false positive (1-specificity) rates of CP, was evaluated from ROC curves at the three pre-defined time points.

Calibration was assessed using the goodness of fit of the prediction models to evaluate the agreement between the observed and the predicted CP in the full sample by Pearson $\chi 2$ test and across different risk deciles by the Hosmer-Lemeshow method (150). Evidence of poor fit is indicated if either test showed statistical significance ($p < 0.05$). Using Pearson $\chi 2$, the training full model fits reasonably well with $p = 0.32$. However, the number of covariate patterns was close to the number of observations (528 and 533 respectively), making the applicability of the Pearson $\chi 2$ test questionable. Therefore, the Hosmer-Lemeshow method regrouped the data by forming 10 almost equal-sized groups based on percentiles of the predicted probabilities of CP (each group has the same or similar predicted probability) (150).

The models' classification was determined using the average classification accuracy (the proportion of correctly classified observations) and the diagnostic properties (sensitivity, specificity, PPV, NPV, and LR) at a selected cut-off of a predicted probability (151). Accuracy was determined as the proportion of both true positive and true negative predicted cases in relation to the whole prediction (both true and false).

The selected cut-off for each model was based on maximizing sensitivity, specificity and correct classification, while considering the clinical context in which the prediction was used. For prenatal and perinatal prediction, the aim was to ensure high specificity and negative predictive value to enable prediction of true negative cases of CP to guide counseling parents when providing life support or intensive care after extreme preterm birth. By contrast, the postnatal prediction is used to counsel parents about the long-term outcome and to refer high risk infants for early intervention, and therefore should have a high sensitivity to identify most cases of CP, but with low false positive rate to avoid creating anxiety and burden on families and overuse of health care services.

Internal validation. The logistic regression models developed on the training dataset were validated using the testing dataset. The AUC of the validated models, their accuracy and diagnostic properties (sensitivity, specificity, PPV, NPV, and LR) were determined at the selected cut-off of a predicted probability.

Objective 2: Random Forest Model

Model building. The RF ensemble method was used to predict CP from prenatal, perinatal and postnatal candidate predictors, as described before. Since RF, in the presence of an imbalance of the predicted classes (11% CP vs. 89% non-CP) will favor the majority class, up-weighting of the minority class (CP) was performed before the analysis to achieve an even ratio of the two classes (152). The number of variables available for splitting at each node (2-10) was optimized using 10-fold cross validation, repeated 10 times, over a parameter grid in the training data; 500 trees were used for the RF. Classification of CP (yes/no) for each observation was based on the majority vote of the RF trees. I qualitatively assessed the relative importance of predictors using the variable importance plots. RF models were implemented in R / RStudio (153) with the *caret* package (154) and the *randomForest* package (155).

Model testing. Due to the upweighting of the minority class, assessment of model calibration was not meaningful. Discrimination was assessed using AUC and the classification accuracy was evaluated based on the confusion matrix for each model.

Internal validation. The RF models developed on the training dataset were validated using the testing dataset. The validated models' AUC, accuracy and diagnostic properties (sensitivity, specificity, PPV, NPV, and LR) were determined.

Objective 3: Comparing Logistic Regression and RF Prediction Models

The c-statistics or AUC, for both the logistic regression and RF full models, provided an assessment of their discriminative ability to correctly identify those preterm children who do develop CP from those who do not. Using the AUC, I quantitatively compared both full prediction models to identify the model that provided a better discrimination of CP. Discrimination was considered poor if AUC was 0.5-0.7, fair if AUC was 0.7-0.8, good if AUC was 0.8-0.9 and excellent if AUC was > 0.9.

4.7. Ethics

The research protocol was reviewed by the Data Management Committee of the PFUP and the Research Ethics Board at the IWK Health Centre (file # 1024274). De-identified data were stored on the NS Health secure network drive and accessed through a password-protected computer at the IWK Health Centre.

CHAPTER 5: RESULTS

5.1 Study Population

A cohort of 1,111 very preterm infants (< 31 weeks' gestation) were born during the study period. After the exclusion of infants with congenital anomalies (n=44) or receiving palliation at delivery (n=117), 950 very preterm infants were eligible for inclusion in this study. Of those, 112 infants died prior to 36 months of corrected age (majority during NICU stay), 16 children were lost to follow-up, and 42 children had their last assessment done prior to 18 months corrected age. A further three infants with influential values were excluded from the analysis at the modeling stage; 1 with hospitalization for 729 days and 2 with > 6 months of mechanical ventilation (4356 hours), leaving 777 preterm survivors (93% of the eligible cohort) in the analysis sample.

The mean gestational age in this cohort was 28 weeks (SD 1.9), ranging from 22 to 30 weeks, and the mean birth weight was 1140 g (SD 324), ranging from 460 to 2180 g. Of the 777 infants included in the analysis, 108 (14%) were of extreme low gestational age (<26 weeks) and 274 (35%) were of extreme low birth weight < 1000 g. The mean duration of follow-up was 35 (SD 9) months of post-term age.

Within this cohort, 86 children were diagnosed with CP (11%) and 691 children were free from CP (89%) (Figure 5.1). Almost two thirds of children with CP (65%) had mild disease (GMFCS level I: n=57) and the majority (84%) were ambulatory (GMFCS level I: n=57, GMFCS level II: n=15). Only 14 children (16%) had non-ambulatory CP (GMFCS level III: n=8, GMFCS level IV: n=4, GMFCS level V: n=2).

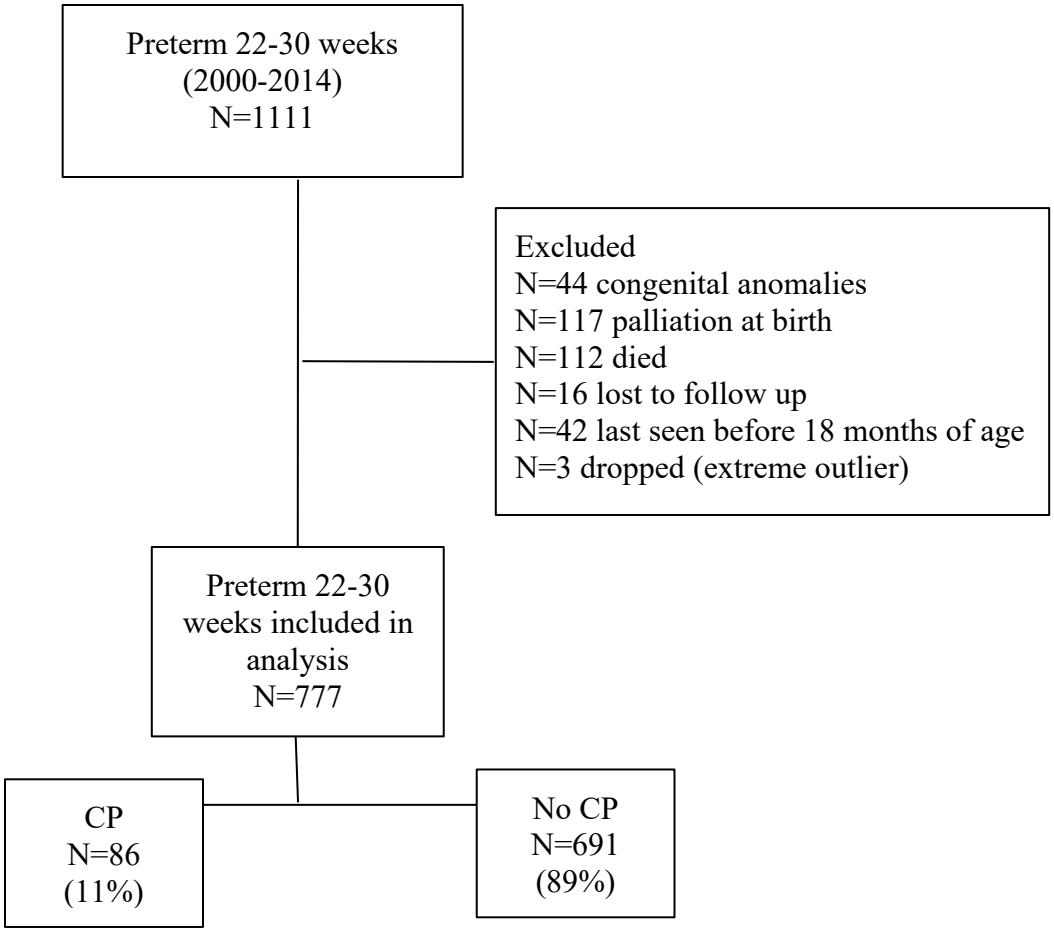

Figure 5.1 Study Population Flow Chart

5.1.1 Comparison of Children With and Without CP in the Study Cohort

Table 5.1 shows the prenatal, perinatal, and postnatal characteristics of children with and without CP.

Prenatal variables: Mothers of children with CP were more likely to be single-parent, but less likely to have suffered from gestational hypertension compared to mothers of children without CP.

Perinatal variables: Children with CP were less likely to be exposed in-utero to antenatal steroids or magnesium sulfate, had lower gestational ages and birth weights compared to those without CP. They were more likely to be of extremely low gestational age and extremely low birth weight compared to those without CP. They also had higher rates of severe birth depression and delivery room resuscitation (chest compression or epinephrine), lower 5-minute Apgar scores, and lower blood pressure and hemoglobin levels within the first 6 hours of admission to NICU compared to children without CP.

Postnatal variables: Children with CP were generally sicker than those without CP as indicated by higher rates of respiratory morbidity, severe brain injury (including cystic white matter lesions, posthemorrhagic hydrocephalus and cerebral parenchymal echodense lesions), hemodynamically significant ductus arteriosus, sepsis, NEC, severe ROP, and hematologic abnormalities compared to those without CP. They also received more intensive medical and surgical therapy during their NICU stay and they had longer median length of hospitalization compared to those without CP (Table 5.1).

Table 5.1: Comparison of Children With and Without CP in the Study Cohort

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
PRENATAL VARIABLES				
Maternal age (mean, SD)	Continuous	-	27.2 (2.1)	27.9 (1.9)
Married or common law	Binary	1	63/86 (73)	584/681 (86)
SES	Categorical	11		
Class I			8/76 (11)	58/613 (10)
Class II			18/76 (24)	209/613 (34)
Class III			22/76 (29)	176/613 (29)
Class IV			15/76 (20)	108/613 (18)
Class V			13/76 (17)	62/613 (10)
Urban residence	Binary	-	67/86 (78)	500/691 (72)
Primigravida	Binary	-	39/86 (45)	290/691 (42)
Multiparity	Binary	-	13/86 (15)	98/691 (14)
Abortion/miscarriages	Binary	-	27/86 (31)	233/691 (34)
Previous stillbirth	Binary	-	<5/86 (5)	16/691 (2)
Smoking	Binary	5	24/76 (32)	180/661 (27)
Substance use	Binary	-	6/86 (7)	76/691 (11)
Psychiatric disease	Binary	-	11/86 (13)	90/691 (13)
Antidepressants	Binary	-	5/86 (6)	50/691 (7)

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
Hypertension	Binary	-	11/86 (13)	141/691 (20)
Diabetes	Binary	-	5/86 (6)	34/691 (5)
Tocolytics	Binary	-	25/86 (29)	247/691 (36)
Indomethacin (for tocolysis)	Binary	-	12/86 (14)	98/691 (14)
Prelabor premature rupture of membranes	Binary	-	16/86 (19)	187/691 (27)
Chorioamnionitis or funisitis	Binary	-	11/86 (13)	84/691 (12)
GBS colonization	Binary	-	8/86 (9)	73/691 (11)
Intrapartum antibiotics	Binary	-	36/86 (42)	309/691 (45)
Abruption	Binary	-	9/86 (11)	49/691 (7)
Antepartum hemorrhage	Binary	-	25/86 (29)	151/691 (22)
Multiples	Binary	-	24/86 (28)	216/691 (31)
Fetal growth restriction	Binary	-	9/86 (11)	97/691 (14)
Fetal distress	Binary	-	8/86 (9)	63/691 (9)
PERINATAL VARIABLES				
Intrapartum magnesium sulfate	Binary	-	13/86 (15)	213/691 (31)

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
Any antenatal steroids	Categorical	0.1	68/86 (79)	636/691 (92)
Antenatal steroids (>24 hours prior to delivery)	Binary	0.1	35/86 (41)	509/690 (74)
Cesarean section	Binary	-	42/86 (49)	408/690 (59)
Moderate-severe asphyxia	Binary	-	75/86 (87)	501/691 (73)
Chest compression/epinephrine	Binary	0.5	15/84 (18)	33/689 (5)
Outborn	Binary	-	9/86 (10)	44/691 (6)
Male sex	Binary	-	42/86 (49)	369/691 (53)
Gestational age in weeks (mean, SD)	Continuous	-	27.2 (2.1)	27.9 (1.9)
Extremely low gestational age	Binary	-	24/86 (28)	84/691 (12)
Birth weight in grams (mean, SD)	Continuous	-	1089 (344)	1148 (321)
Extremely low birth weight	Binary	-	39/86 (45)	235/691 (34)
Birth weight z-score (mean, SD)	Continuous	0.2	0.125 (0.88)	-0.035 (0.84)
Small for gestational age	Binary	-	6/86 (7)	49/691 (7)
Apgar at 1 minute (median, IQR)	Continuous	1	4 (4)	5 (3)
Apgar at 5 minutes (median, IQR)	Continuous	1	7 (2)	8 (3)
Admission temperature (mean, SD)	Continuous	2	36.6 (0.8)	36.7 (0.8)
Admission hemoglobin	Continuous	1	153 (28)	162 (27)

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
(mean, SD)				
Hypotension on NICU admission	Binary	2	20/83 (24)	94/679 (14)
POSTNATAL VARIABLES				
Lowest hemoglobin during first 24 hours (median, IQR)	Continuous	0.6	83 (12)	86 (6)
Severe IVH (≥ grade 3 IVH)	Binary	-	35/86 (41)	38/691 (6)
Posthemorrhagic hydrocephalus	Binary	-	22/86 (26)	11/691 (2)
Ventriculoperitoneal shunt	Binary	-	18/86 (21)	5/691 (1)
Cystic brain lesions (PVL, porencephaly)	Binary	-	40/86 (47)	9/691 (1)
Parenchymal echodense brain lesions	Binary	-	15/86 (17)	99/691 (14)
Severe RDS	Binary	-	80/86 (93)	453/691 (66)
Surfactant for RDS	Binary	-	77/86 (90)	486/691 (70)
Hours on mechanical ventilation (median, IQR)	Continuous	-	349 (1091)	44 (442.3)
Cystic BPD	Binary	-	28/86 (33)	119/691 (17)
Dexamethasone for BPD	Binary	-	26/86 (30)	107/691 (16)
Home oxygen at discharge	Binary	-	<5/86 (4)	30/691 (4)
Nasal CPAP	Binary	-	63/86 (73)	552/691 (80)
High frequency oscillation	Binary	-	22/86 (26)	65/691 (9)

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
Pneumothorax	Binary	-	<5/86 (5)	30/691 (4)
Pulmonary hemorrhage	Binary	-	5/86 (6)	18/691 (3)
Treatment for ROP	Binary	-	50/86 (58)	320/691 (46)
Significant PDA	Binary	-	37/86 (43)	167/691 (24)
Medical treatment for PDA	Binary	-	40/86 (47)	212/691 (31)
PDA ligation	Binary	-	18/86 (21)	60/691 (9)
NEC ≥ Bell stage 2	Binary	-	6/86 (7)	18/691 (3)
Duration of TPN (Median, IQR)	Continuous	2	40 (34)	26 (27)
Neonatal septicemia	Binary	-	32/86 (37)	160/691 (23)
Clinical (culture negative) sepsis	Binary	-	5/86 (6)	72/691 (10)
Systemic infection	Binary	-	38/86 (44)	242/691 (35)
Neonatal anemia	Binary	-	77/86 (90)	522/691 (76)
Neonatal thrombocytopenia (<100,000)	Binary	-	33/86 (38)	123/691 (18)
Severe neonatal hypoglycemia (<1.67 mmol/L)	Binary	-	14/86 (16)	77/691 (11)
Insulin for neonatal hyperglycemia	Binary	-	22/86 (26)	70/691 (10)
Inhaled nitric oxide	Binary	-	13/86 (15)	37/691 (5)
Prophylactic indomethacin	Binary	0.1	9/85 (11)	68/691 (10)
Inotropes	Binary	-	34/86 (40)	79/691 (11)
Muscle relaxant	Binary	-	26/86 (30)	111/691 (16)

Variable	Type	% Missing	CP (n=86)	No CP (n=691)
Major surgery	Binary	-	21/86 (24)	53/691 (8)
Antireflux medications	Binary	-	14/86 (16)	97/691 (14)
Resuscitation during NICU stay	Binary	-	13/86 (15)	32/691 (5)
Hospitalization days (median, IQR)	Continuous	1	87 (60)	70 (44)

Data are presented as n/N (percentage), mean (±SD) or median (IQR). The dash indicates no missing data
Abbreviations: BPD (bronchopulmonary dysplasia), CP (cerebral palsy), CPAP (continuous positive airway pressure), GBS (group B streptococci), IVH (intraventricular hemorrhage), IQR (interquartile range), NEC (necrotizing enterocolitis), NICU (Neonatal Intensive Care Unit), PDA (patent ductus arteriosus), PVL (periventricular leukomalacia), RDS (respiratory distress syndrome), ROP (retinopathy of prematurity), SD (standard deviation), SES (socioeconomic status).

The full population dataset was then randomly assigned to a training set (70%, n=544) to develop the prediction models and a testing set (30%, n=233) to test their predictive performance.

Table 5.2: Random Splitting of Population Dataset

	CP	No CP	Total
Testing Dataset	26 (11)	207 (89)	233 (100)
Training Dataset	60 (11)	484 (89)	544 (100)
Total	86 (11)	691 (89)	777 (100)

Data are presented as number (percentage)

5.2 Logistic Regression Model Development and Testing

The development and internal validation of the CP prediction models at the three time points (prenatal, perinatal, and postnatal) was reported as per the TRIPOD statement "Transparent Reporting of a Multivariable prediction model for Individual Prognosis or Diagnosis" (TRIPOD) [http://www.tripod-statement.org/TRIPOD/TRIPOD-Checklists].

5.2.1 Univariate analysis of the Training Dataset

The list of the prenatal, perinatal, and postnatal candidate predictors and their unadjusted association with CP (OR with 95% CI) are provided in Tables 5.3 (1-3). Thirty-eight candidate predictors were associated with CP at the three time points with $p < 0.05$.

Table 5.3.1 Prenatal Risk Factors Associated with CP on Univariate Analysis

Variable		Unadjusted OR	(95% CI)
Maternal age		0.97	(0.93-1.02)
Married or common law		**0.4**	**(0.22-0.74)**
SES	Class I	Ref.	Ref.
	Class II	0.59	(0.19-1.80)
	Class III	0.80	(0.27-2.38)
	Class IV	1.41	(0.65-6.41)
	Class V	2.04	(0.47-4.24)
Urban residence		**2.28**	**(1.09-4.76)**
Primigravida		1.01	(0.58-1.73)
Multiparity		1.12	(0.53-2.38)
Abortion/miscarriage		1.03	(0.59-1.81)
Previous stillbirths		3.13	(0.81-12.14)
Smoking		1.21	(0.66-2.24)
Substance use		0.79	(0.30-2.06)
Psychiatric disease		0.72	(0.30-1.73)
Antidepressants		0.92	(0.31-2.67)
Hypertension		0.51	(0.22-1.15)
Diabetes		1.43	(0.48-4.29)
Tocolytics		0.79	(0.44-1.43)
Indomethacin		1.56	(0.77-3.16)
Prelabor premature rupture of membranes		0.67	(0.35-1.31)
Chorioamnionitis or funisitis		1.45	(0.72-2.93)
Maternal antibiotics		0.96	(0.56-1.66)
Antepartum hemorrhage		**1.90**	**(1.07-3.36)**

Variable	Unadjusted OR	(95% CI)
Multiples	0.83	(0.46-1.53)
Fetal growth restriction	0.84	(0.37-1.92)
Fetal distress	1.17	(0.48-2.88)

Data presented as unadjusted OR (95% CI), bold font indicates statistical significance

Abbreviations: CI (confidence interval), CP (cerebral palsy), OR (Odds ratio), SES (socioeconomic status)

Table 5.3.2: Perinatal Risk Factors Associated with CP on Univariate Analysis

Variable	Unadjusted OR	(95% CI)
Cesarean Section	**0.55**	**(0.32-0.94)**
Intrapartum magnesium sulfate	**0.40**	**(0.19-0.84)**
Antenatal steroids	**0.23**	**(0.13-0.40)**
Moderate-severe asphyxia	**2.49**	**(1.15-5.38)**
Chest compression/epinephrine	**4.03**	**(1.87-8.66)**
Gestational age	**0.83**	**(0.73-0.95)**
Extreme prematurity (< 26 weeks)	**2.92**	**(1.58-5.34)**
Birth weight	1.0	(1.00-1.00)
Birth weight z score	1.34	(0.97-1.85)
Extreme low birth weight (<1000 grams)	1.48	(0.86-2.55)
Small for gestational age (<10th centile)	1.28	(0.48-3.43)
Male	0.79	(0.46-1.36)
Apgar at 1 minute	**0.82**	**(0.73-0.92)**
Apgar at 5 minutes	**0.77**	**(0.67-0.88)**
Admission temperature	0.85	(0.57-1.27)
Admission hemoglobin	0.99	(0.98-1.00)
Hypotension on admission to NICU	**2.09**	**(1.08-4.04)**
Outborn	1.47	(0.59-3.66)

Data presented as unadjusted OR (95% CI), bold font indicates statistical significance

Abbreviations: CI (confidence interval), CP (cerebral palsy), NICU (Neonatal Intensive Care Unit), OR (Odds ratio)

Table 5.3.3: Postnatal Risk Factors Associated with CP on Univariate Analysis

Variable	Unadjusted OR	(95% CI)
Lowest hemoglobin during first 24 hours	**0.97**	**(0.96-0.99)**
Severe IVH (≥ grade 3 IVH)	**10.46**	**(5.52-19.81)**
Posthemorrhagic hydrocephalus	**19.83**	**(7.98-49.32)**
Ventriculoperitoneal shunt	**40.08**	**(10.93-147.00)**

Variable	Unadjusted OR	(95% CI)
Cystic brain lesions (PVL, porencephaly)	**83.83**	**(30.33-231.68)**
Parenchymal echodense brain lesions	1.24	(0.60-2.57)
Severe RDS	**5.00**	**(2.11-11.88)**
Surfactant for RDS	**3.21**	**(1.42-7.22)**
Nasal CPAP	0.63	(0.34-1.16)
Home oxygen at discharge	1.11	(0.32-3.81)
Duration of tracheal intubation	1.00	(1.00-1.00)
Cystic BPD	**2.42**	**(1.36-4.32)**
Dexamethasone	**1.86**	**(1.01-3.41)**
Pneumothorax	0.72	(0.17-3.16)
Pulmonary hemorrhage	2.26	(0.61-8.35)
NEC (≥ stage 2 Bell's)	**3.91**	**(1.31-11.67)**
Days of parenteral nutrition	1.02	(1.00-1.03)
Neonatal septicemia	**2.32**	**(1.33-4.06)**
Clinical sepsis	0.66	(0.23-1.91)
Systemic infection	**1.96**	**(1.15-3.30)**
Severe ROP (≥ stage 3)	**3.70**	**(1.90-7.22)**
Treatment for ROP	1.44	(0.84-2.48)
Prophylactic Indomethacin	0.50	(0.15-1.65)
Significant PDA	**2.60**	**(1.50-4.49)**
Treatment for PDA	**1.93**	**(1.12-3.32)**
PDA ligation	**2.97**	**(1.52-5.82)**
Neonatal thrombocytopenia (<100,000 10 e/L))	**3.14**	**(1.77-5.57)**
Neonatal anemia	**3.55**	**(1.39-9.07)**
Severe neonatal hypoglycemia	1.63	(0.78-3.40)
Inhaled nitric oxide	**3.52**	**(1.66-7.49)**
Antireflux medications	0.84	(0.38-1.84)
Insulin for hyperglycemia	**3.46**	**(1.81-6.62)**
High frequency oscillatory ventilation	**3.30**	**(1.73-6.30)**
Inotropes	**4.48**	**(2.49-8.06)**

Variable	Unadjusted OR	(95% CI)
Muscle paralysis	**2.01**	**(1.10-3.66)**
Major surgery	**1.26**	**(1.90-7.22)**
Resuscitation during NICU stay	**3.24**	**(1.43-7.32)**
Length of hospitalization	1.00	(1.00-1.01)

Data presented as unadjusted OR (95% CI), bold font indicates statistical significance

Abbreviations: BPD (bronchopulmonary dysplasia), CI (confidence interval), CP (cerebral palsy), CPAP (continuous positive airway pressure), IVH (intraventricular hemorrhage), NEC (necrotizing enterocolitis), NICU (Neonatal Intensive Care Unit), OR (Odds ratio), PDA (patent ductus arteriosus), PVL (periventricular leukomalacia), RDS (respiratory distress syndrome),ROP (retinopathy of prematurity)

5.2.2 Multivariable Logistic regression Model Development on Training Dataset

The prediction models at each time point with their predictors and diagnostic properties are reported below. Table 5.4 compares the individual logistic regression and the full model developed from these multiple regressions in incremental steps using prenatal, perinatal, and postnatal predictors. The predictors for each model with their adjusted OR (95% CI) and the model's AUC are provided.

Prenatal Model (Time Point 1)

Maternal characteristics identified as independent predictors of CP in the Prenatal model included: marital status (being married or common-law), receipt of tocolytics, hypertension, geographic area of residence, and indomethacin therapy). No evidence of a poor fit was demonstrated by the goodness of fit ($p = 0.95$) and Hosmer-Lemeshow ($p = 0.66$) tests. The Prenatal model discrimination was poor (AUC 0.68, 95% CI 0.61-0.76). The model correctly classified 80% of children with respect to CP status; the model's diagnostic properties were: sensitivity 42%, specificity 85%, PPV 26%, NPV 92%, LR+ 2.75, LR- 0.69, false positive rate 15%, and false negative rate 58%.

Perinatal Model

Using only perinatal predictors, antenatal steroids, small for gestational age, receipt of resuscitation at birth (chest compression or epinephrine) and one unit change in the exponentially transformed birth weight z scores (cubed birth weight z scores) were predictors of CP, when other predictors were kept constant. There was no evidence of poor fit as demonstrated by the goodness of fit ($p = 0.68$) and Hosmer-Lemeshow ($p = 0.62$) tests. The perinatal model discrimination was fair (AUC of 0.76, 95 % CI 0.69-0.83).

Combined Prenatal and Perinatal Model (Time Point 2)

In the combined prenatal/perinatal model, marital status, receipt of antenatal steroids, small for gestational age, resuscitation at birth, and birth weight z-score were predictors of CP. There was no evidence of poor fit as demonstrated by testing for goodness of fit ($p = 0.2$) and Hosmer-Lemeshow ($p = 0.7$) tests. Adding perinatal variables to the prenatal model improved the discriminative performance from poor to fair (AUC 0.77, 95% CI 0.70-0.84). The combined prenatal-perinatal model correctly classified 77% of children; the model's diagnostic properties were: sensitivity 68%, specificity 78%, PPV 27%, NPV 95%, LR+ 3.06, LR- 0.41.

Postnatal Model

Cystic white matter lesions, brain parenchymal echodensities, posthemorrhagic hydrocephalus, severe RDS, treatment for ROP, receipt of inotropes, and receipt of nasal CPAP were predictors of CP. There was no evidence of poor fit as demonstrated by goodness of fit ($p = 0.99$) and Hosmer-Lemeshow ($p = 0.84$) tests. The postnatal model discrimination was good (AUC 0.88 (95 % CI 0.83-0.93).

Full Model (Combined Prenatal/Perinatal/Postnatal Model) (Time point 3)

All prenatal, perinatal, and postnatal predictors from the three models were combined to develop the full prediction model of CP. Marital status, antenatal steroids, birth weight z-score, cystic white matter lesions, brain parenchymal echodensities, posthemorrhagic hydrocephalus, treatment for ROP, neonatal thrombocytopenia, receipt of nasal CPAP and medical treatment of reflux were independent predictors of CP. There was no evidence of

poor fit by goodness of fit (p = 0.32) and Hosmer-Lemeshow (p = 0.53) tests. The full model had excellent discrimination with AUC of 0.91 (95% CI 0.87-0.96) and correctly classified 86% of children with and without CP. The model diagnostic properties were: sensitivity 80%, specificity 87%, PPV 43%, NPV 97%, LR+ 6.29, LR- 0.23.

5.2.3 Comparison of Logistic Regression Models of CP on the Training Dataset

The performance of the logistic regression models in the training dataset at the three time points (prenatal, perinatal, and postnatal) was assessed using discrimination, calibration and classification. Table 5.4 compares the logistic regression models of CP in the training dataset.

Table 5.4: Logistic Regression Models of CP in the Training Dataset

	Prenatal Model (Time point 1)	**Perinatal Model**	**Prenatal/ Perinatal Model (Time point 2)**	**Postnatal Model**	**Full Model (Time point 3)**
No. candidate predictors	25	19	13	38	24
No. predictors in model	5	4	5	7	11
AUC (95% CI)	0.68 (0.61-0.76)	0.76 (0.69-0.83)	0.77 (070-0.84)	0.88 (0.83-0.93)	0.91 (0.87-0.96)
Odds ratio (95% CI)					
Prenatal Variables					
Married or common law	0.44 (0.23-0.85)		0.48 (0.24-0.98)		0.24 (0.11-0.55)
Urban residency	2.64 (1.10-4.98		2.07 (0.95-4.54)		
Maternal Indomethacin	3.21 (1.09-9.50)				
Tocolytics	0.39				

	Prenatal Model (Time point 1)	**Perinatal Model**	**Prenatal/ Perinatal Model (Time point 2)**	**Postnatal Model**	**Full Model (Time point 3)**
	(0.16-0.98)				
Maternal hypertension	0.37 (0.15-0.91)				
Maternal diabetes	2.64 (0.83-8.42)				
Smoking	1.01 (0.99-1.02)				
Prelabor premature rupture of membranes	0.52 (0.26-1.03)				
Perinatal Variables					
Antenatal steroids		0.21 (0.11-0.40)	0.22 (0.12-0.42)		0.43 (0.20-0.93)
Resuscitation at birth		2.53 (1.07-5.97)	3.38 (1.46-7.86)		
Small for gestational age		6.21 (1.67-23.05)	5.51 (1.47-20.67)		
Birth weight z-score		1.75 (1.16-2.63)	1.73 (1.15-2.61)		1.81 (1.13-2.89)
Hypotension on admission		2.02 (0.98-4.15)	1.87 (0.90-3.87)		
Postnatal Variables					

	Prenatal Model (Time point 1)	Perinatal Model	Prenatal/ Perinatal Model (Time point 2)	Postnatal Model	Full Model (Time point 3)
Cystic white matter lesions				128.63 (34.80-475.39)	100.57 (29.70-340.61)
Parenchymal echodensities				2.76 (1.09-6.81)	3.45 (1.38-8.63)
Posthemorrhagic hydrocephalus				7.76 (1.93-31.20)	5.36 (1.32-21.84)
Severe RDS				3.99 (1.36-11.72)	
Nasal CPAP				0.17 (0.06-0.48)	0.34 (0.14-0.84)
Inotropes				3.91 (1.25-12.19)	
ROP treatment				3.97 (1.62-9.73)	2.68 (1.16-6.16)
Thrombocytopenia				2.48 (0.98-6.24)	2.60 (1.01-6.68)
Reflux treatment				0.30 (0.09-1.01)	0.26 (0.08-0.91)
Necrotizing enterocolitis				0.05)0.02-1.53)	
Dexamethasone				0.36 (0.11-1.15)	

Abbreviations: AUC (Area under the curve), CI (confidence interval), CP (cerebral palsy), CPAP (continuous positive airway pressure, OR (Odds ratio), RDS (respiratory distress syndrome), ROP (retinopathy of prematurity).

Discrimination of the three logistic regression models of CP in the training dataset was determined by AUC as shown from ROC curves in Figure 5.2. The Prenatal model had poor discrimination of CP. The model's discrimination was fair when perinatal predictors were added to the prenatal model. The combined prenatal/perinatal/postnatal predictors had excellent discrimination with an AUC of 0.91 (95% CI 0.87-0.96).

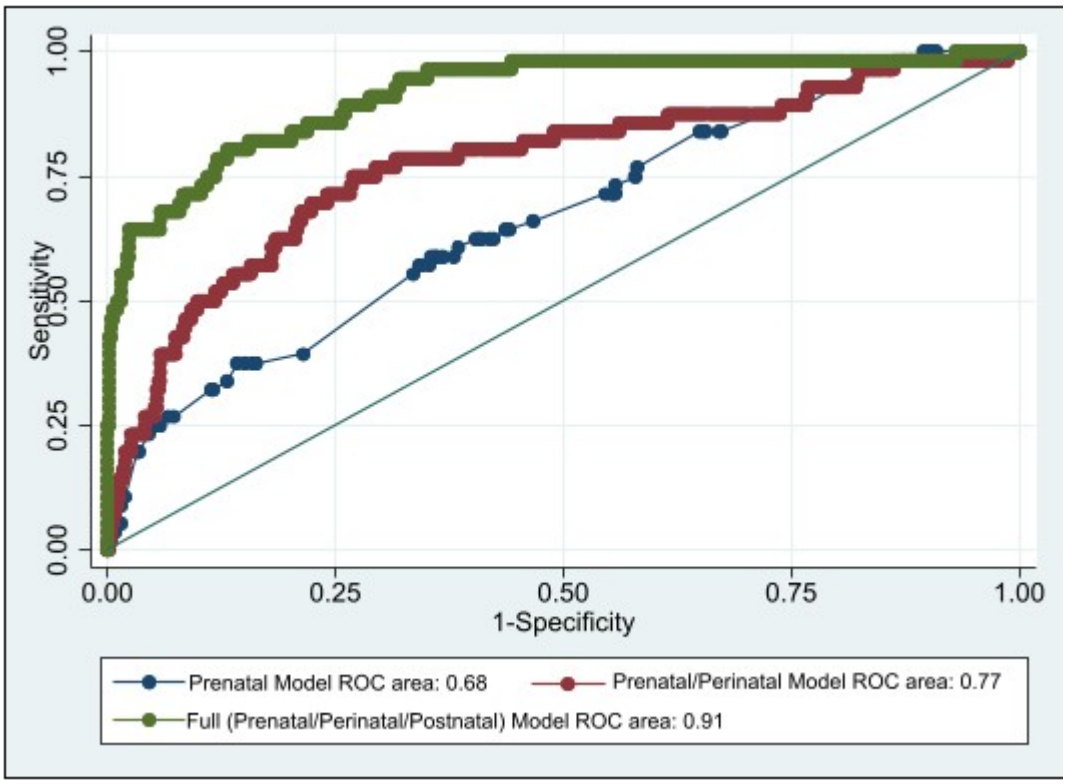

Figure 5.2 The ROC Curves of Logistic Regression Models of CP in the Training Dataset

The calibration for all the three logistic regression prediction models in the training dataset showed no evidence of poor fit as demonstrated by goodness of fit and Hosmer-Lemeshow tests.

The classification of the three models in the training dataset was determined by the proportion of correctly classified observations at the selected cut-off points, as described before. Both the prenatal and the combined prenatal-perinatal models had poor sensitivity (42% and 68%, respectively) but relatively good specificity (85% and 78%, respectively). They both had high NPV (92% and 95%, respectively) and correctly classified 80% and 77% of children with and without CP, respectively. The full model provided the highest sensitivity, specificity and NPV (80%, 87%, 97% respectively) with a low false positive rate of 13% and yielded the best accuracy correctly classifying 86% of children on the basis of CP status.

5.2.4 Internal Validation of the Developed Logistic Regression Model of CP

The AUC of the models in the testing dataset at the three time points are shown in Figure 5.3. Similar to the training dataset, the prenatal model including only maternal and fetal predictors resulted in poor discrimination of CP. Combining prenatal and perinatal predictors improved the model's discrimination only slightly but remained poor. The full model combining predictors from all three time points showed a good discrimination with an AUC of 0.84 (95% CI 0.74-0.95).

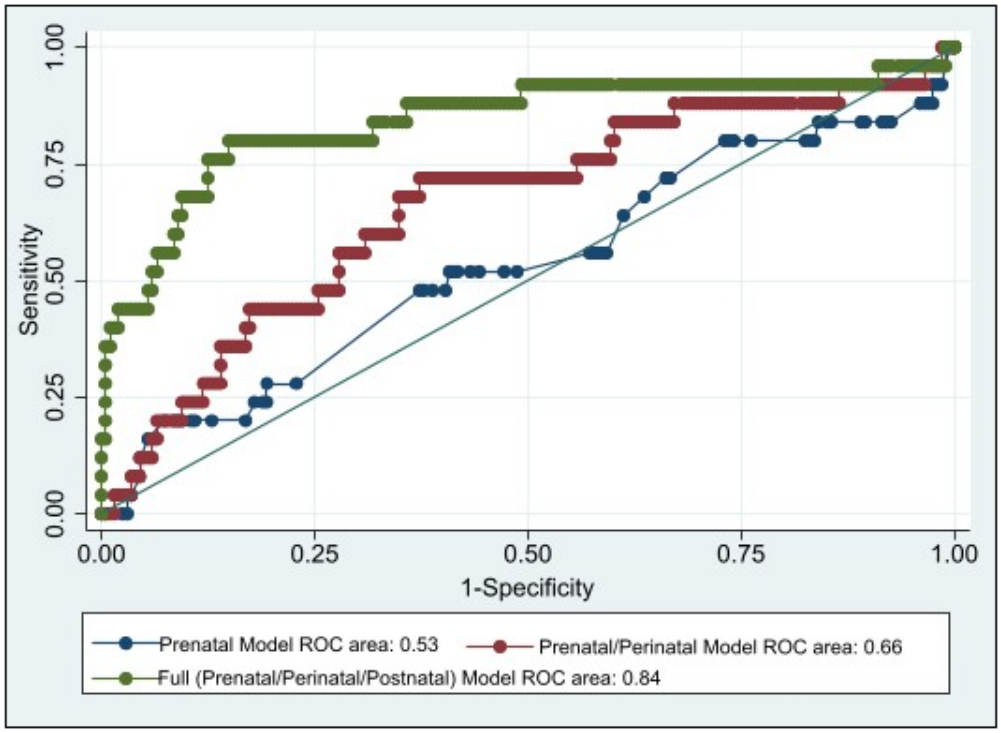

Figure 5.3: The ROC Curves of the Logistic Regression Models of CP in the Testing Dataset

The calibration for all three models of CP in the testing dataset showed no evidence of poor fit as demonstrated by goodness of fit and Hosmer-Lemeshow tests.

Table 5.5 compares the discrimination (AUC), correct classification (accuracy) and diagnostic properties of the three validation models. Both the prenatal and perinatal models had poor sensitivity, but relatively good specificity. They both had high NPV (89% and 92%, respectively) and correctly classified 74% and 76% of children with and without CP, respectively. The full model provided the highest sensitivity, specificity and NPV (77%, 85%, 97% respectively) with a low false positive rate of 15% and yielded the best accuracy correctly classifying 84 % of children with and without CP in this cohort.

5.3 RANDOM FOREST DEVELOPMENT AND TESTING

The discrimination (AUC), accuracy and the diagnostic properties of the RF prediction models of CP at the three time points in the testing dataset are shown in Table 5.5.

Figure 5.4 shows the AUC of the three RF models in the testing dataset. Similar to logistic regression, the prenatal model discrimination was poor. The addition of perinatal predictors to the prenatal model improved the model performance by 11%, resulting in fair discrimination. The full model resulted in 14% further improvement and good discrimination with an AUC of 0.83 (95% CI 0.73-0.93).

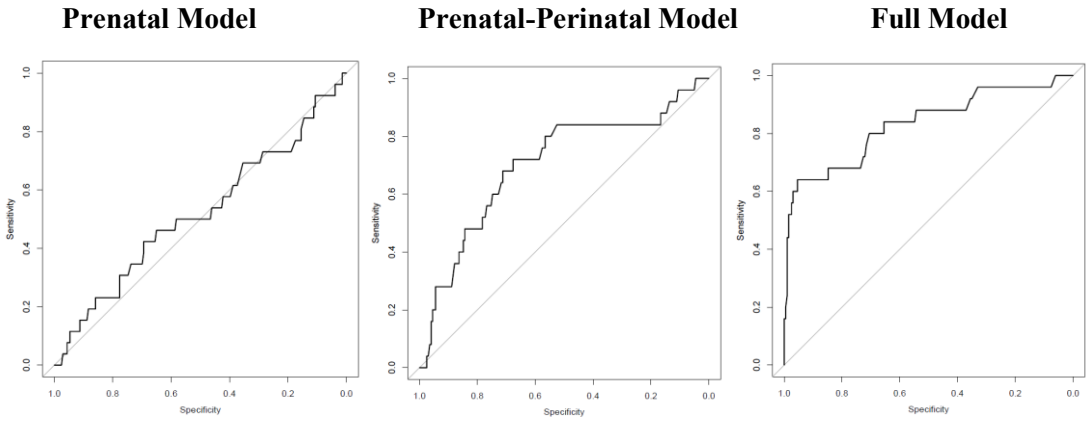

Figure 5.4: The ROC Curves of RF Models of CP in the Testing Dataset

The prenatal and combined prenatal-perinatal RF models correctly classified 72% and 82% of children with and without CP. The full RF prediction model, combining predictors of all three time points, had the best accuracy correctly classifying 91% of children with and without CP in this cohort of very preterm children.

Figures 5.5.1-5.5.3 show the variable importance plots of the three RF models. Maternal age, antenatal steroids and cystic white matter lesions were identified as the most important predictors in the prenatal, combined prenatal-perinatal, and the full model, respectively.

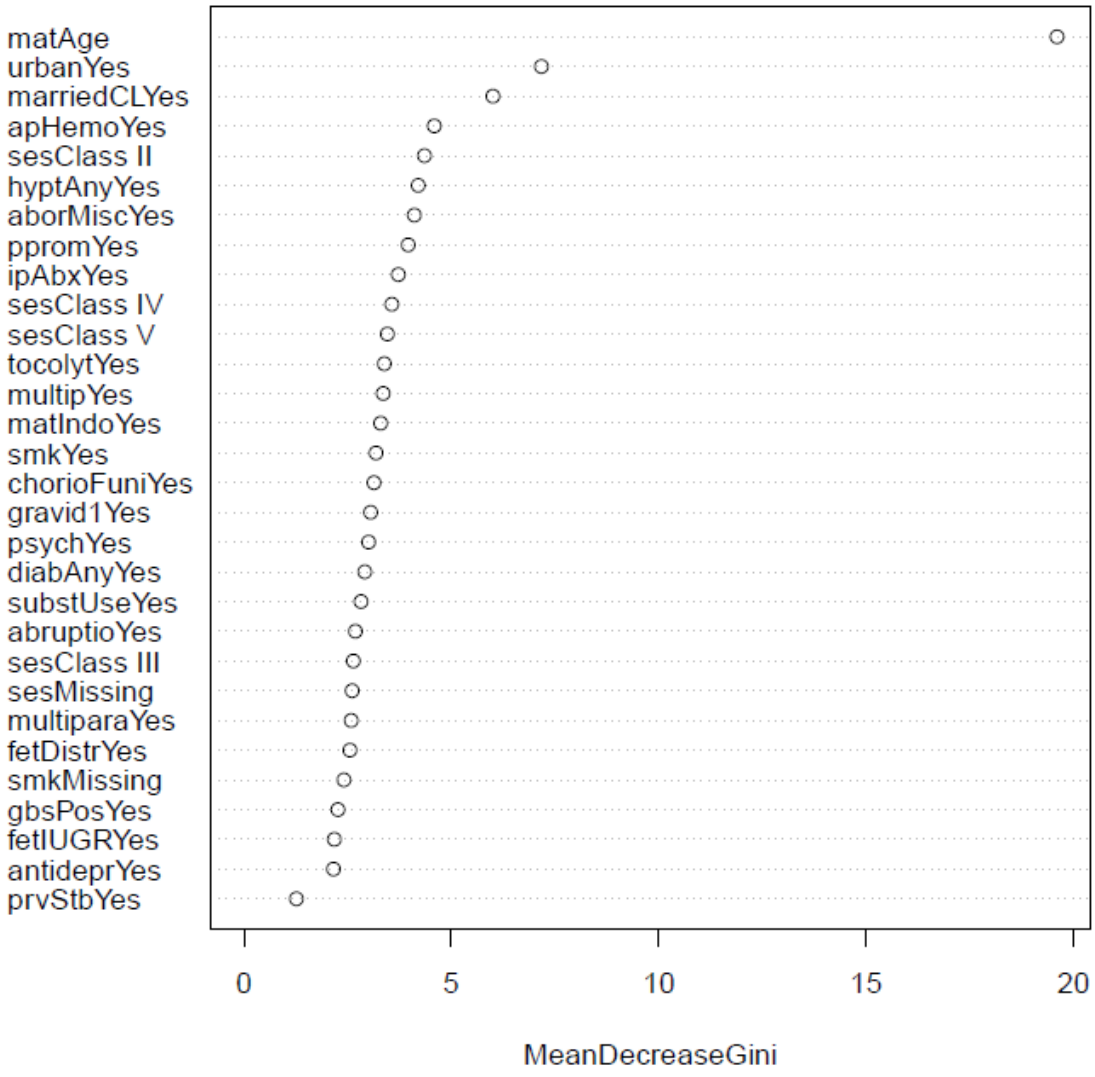

Figure 5.5.1: Variable Importance Plot of the Prenatal RF Model in the Testing Dataset

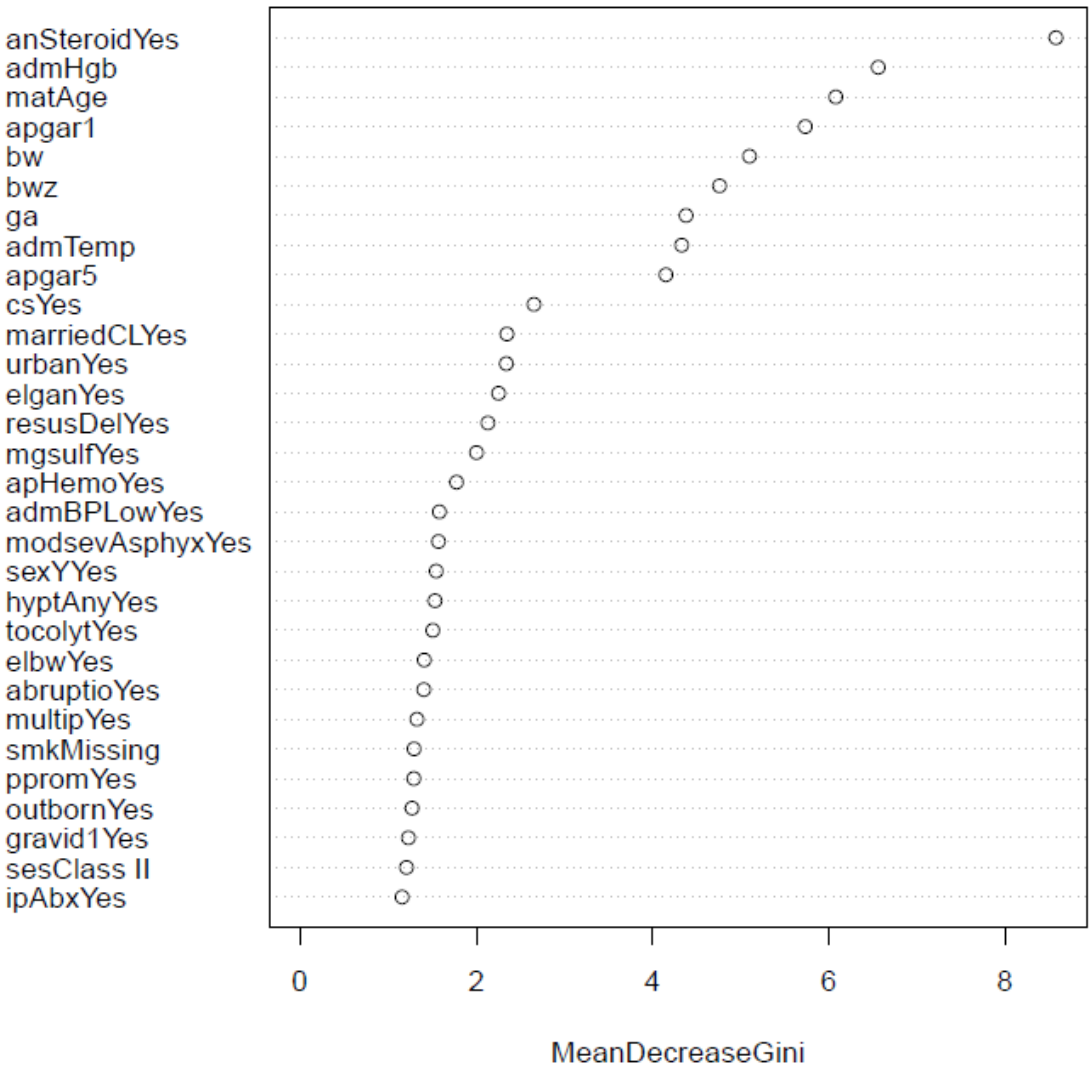

Figure 5.5.2: Variable Importance Plot of the Prenatal-Perinatal RF Model in the Testing Dataset

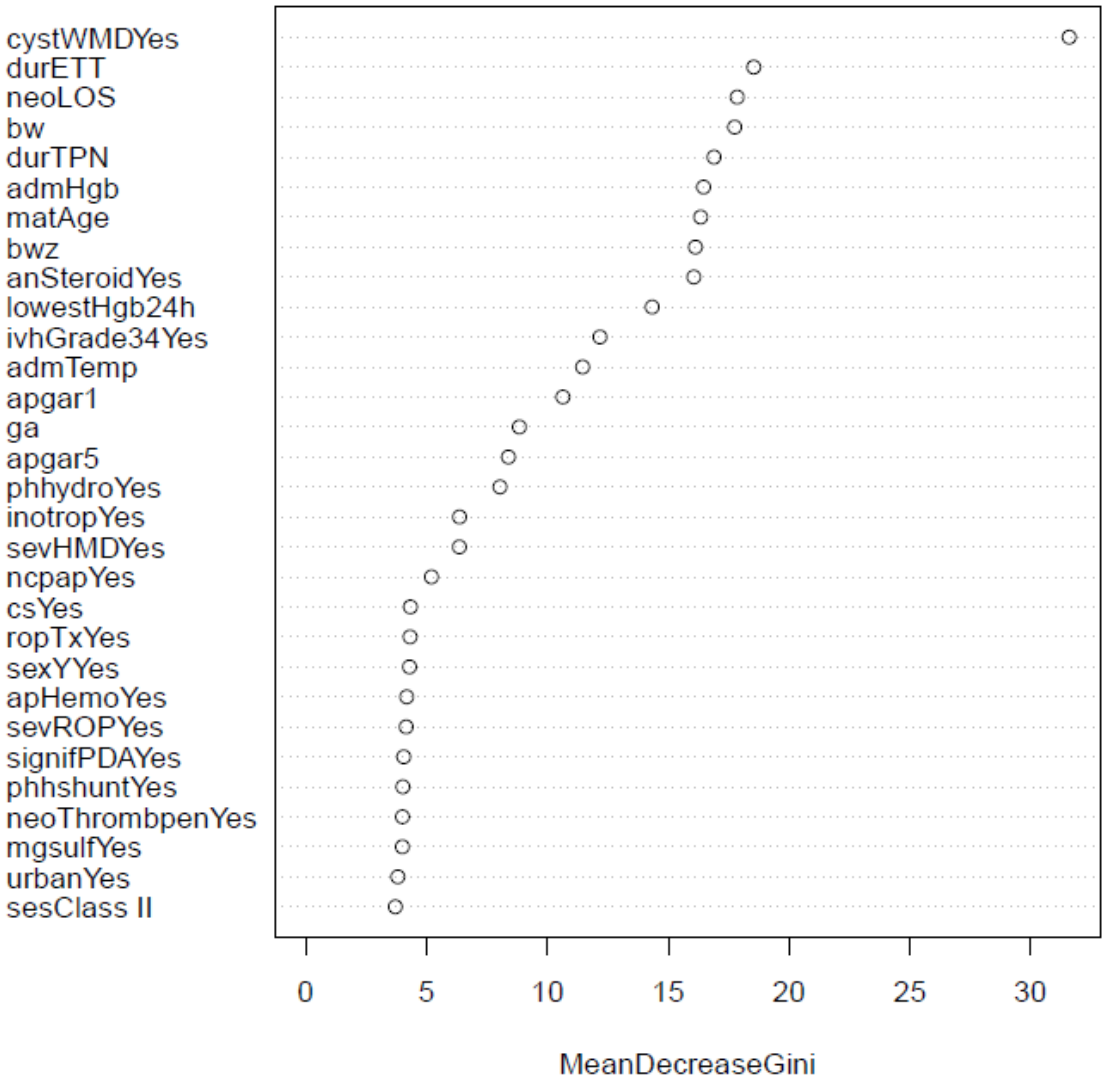

Figure 5.5.3: Variable Importance Plot of the Full RF Model in the Testing Dataset

5.4 Comparing Logistic Regression and RF Prediction Models

Table 5.5 shows the comparison of the discrimination (AUC) and accuracy (correct classification) between the validated logistic regression and RF models in the testing dataset at the three time points. The discrimination (AUC) and accuracy (correct classification) were similar between RF and logistic regression at all three time points. For the full prediction model, both RF and logistic regression yield comparable AUC (0.83 and 0.84, respectively) with similar precision for both methods. Similarly, RF prenatal and perinatal models yield comparable AUC and confidence intervals compared to the corresponding logistic regression models. Regardless of the method used for prediction, including only prenatal predictors resulted in poor discrimination of CP. Combining prenatal and perinatal predictors, slightly improved the discrimination, but it remained only poor to fair. The full prediction model, combining predictors from all three time points, resulted in good discrimination and the best accuracy in prediction of CP.

Table 5.5: Comparison of RF and Logistic Regression Models of CP in the Testing Datasets

	Prenatal Model		**Prenatal/Perinatal Model**		**Full Model**	
	Log. Reg.	RF	Log. Reg.	RF	Log. Reg.	RF
AUC	0.53	0.52	0.66	0.70	0.84	0.83
(95% CI)	(0.41-0.66)	(0.39-0.65)	(0.54-0.78)	(0.58-0.82)	(0.74-0.95)	(0.73-0.93)
Accuracy	74%	72%	76%	82%	84%	91%

Data presented as percentages and AUC (95% CI)

Abbreviations: AUC (area under the curve), CI (confidence interval), Log. Reg. (logistic regression), LR (likelihood ratio), NPV (negative predictive value), PPV (positive predictive value)

Tables 5.6 and 5.7 show the comparison of the diagnostic properties, at the three time points, between the logistic regression models and RF models in the testing dataset, respectively.

Table 5.6: Comparison of the Diagnostic Properties of Logistic Regression Models of CP in the Testing Dataset

	Prenatal Model	Prenatal/Perinatal Model	Full Model
Cutoff	0.16	0.14	0.1
Sensitivity	23%	44%	77%
Specificity	81%	80%	85%
PPV	13%	22%	40%
NPV	89%	92%	97%
LR^+	1.19	2.21	5.26
LR^-	0.95	0.70	0.27

Data presented as percentages

Abbreviations: LR (likelihood ratio), NPV (negative predictive value), PPV (positive predictive value)

Table 5.7: Comparison of the Diagnostic Properties of RF Models of CP in the Testing Dataset

	Prenatal Model	Prenatal/ Perinatal Model	Full Model
Sensitivity	23%	36%	24%
Specificity	79%	88%	98%
PPV	12%	27%	75%
NPV	89%	91%	91%
LR^+	1.00	3.00	12.0
LR^-	1.00	0.73	0.78

Data presented as percentages

Abbreviations: LR (likelihood ratio), NPV (negative predictive value), PPV (positive predictive value)

CHAPTER 6: DISCUSSION

6.1 Summary of Results

In this study, I developed and compared prediction models for CP using logistic regression and RF in a population cohort of very preterm children in Nova Scotia. The full prediction model of CP was developed by incremental addition of predictors that were identified from three models that included prenatal, perinatal, and postnatal variables, respectively. Logistic regression identified maternal marital status (being married or common-law), antenatal steroids, birth weight z scores, cystic white matter lesions, parenchymal echodensities, hydrocephalus, treatment of severe ROP, receipt of nasal CPAP, treatment of reflux and neonatal thrombocytopenia, as independent predictors of CP. By contrast, the variable importance plot from RF ranked maternal age, antenatal steroids and cystic white matter lesions to be the most important predictors of CP at the three time points, respectively.

On internal validation, both RF and logistic regression provided good discrimination between children with and without CP in this cohort. In this study, using clinical predictors, both RF and logistic regression provided similar AUC (0.83 and 0.84, respectively) and comparable classification accuracy (91% and 84%, respectively). Regardless of the method used for prediction, the full model that included predictors from all three time points provided the highest discrimination and the best accuracy.

This is the first population-based study that developed a prediction model for CP based on clinical predictors using the RF ensemble method. The study is also the first to compare the predictive performance of RF to the traditionally used logistic regression in this context.

6.2 Generalizability and Validity

The CP prediction models in this study are likely to be generalizable as they were developed from a sample that is representative to other preterm populations in Canada and developed countries. Those populations share similar characteristics in relation to survival,

CP rates and the maternal and infants' characteristics that are considered to be antecedent to CP, as shown in table 5.1. The 12% mortality rate of eligible infants in this cohort (112/950) is comparable to that reported in the literature, including the International Network for Evaluation of Outcomes (iNEo) in neonates (2, 156-158). The iNEo reported hospital mortality rate of 9.9% (range 4.7%-17%) among 154,233 infants (<32 weeks, <1500 grams) across 11 developed countries including: Canada, Australia, New Zealand, Finland, Sweden, Switzerland, United Kingdom, Tuscany region in Italy, Spain, Israel and Japan (2). In this cohort, 11% of children born very preterm developed CP, which agrees with the incidence rates of CP reported by large population-based studies from Canada and other developed countries (8, 9, 20, 26, 28). Of note, Center variability and care practices may impact the rates of brain injury antecedent to CP, limiting the generalizability of this analysis. However, the rate of cystic white matter lesions, alone or as part of severe neurologic injury, identified as the most important predictor in this study, is comparable to the published rates in large cohorts of preterm infants (52-54, 157, 158).

A common concern regarding the applicability or the generalizability of prediction models is the inconsistency in the definition, assessment procedure or timing of the predictors. Database studies that span a longer time period generally have the limitation of changes in variable definitions over time or even changes in the diagnostic criteria to define some variables. However, the AC Allen Provincial PFUP database has kept all codes constant with addition of new codes over time and the program director (M. Vincer) and database manager conduct periodic audits to ensure the accuracy and the reliability of the database coding. Importantly, the definition of CP has been standardized in the database since its inception and the severity classification of CP using GMFCS has been standardized in the database since 2000 (over the study period).

In diagnostic prediction models, bias may be introduced by misclassification of outcome status due to multiple definitions of the outcome, assessment done at different times, assessment not following a standardized referenced method or requiring subjective interpretation. The risk of ascertainment bias is low in this study, since the clinical diagnosis of CP relied on standardized multiple neuromotor assessment of all infants in

this cohort at regular intervals, regardless of their postnatal course of illness. The ascertainment of the outcome in this study was done using a standardized validated method (GMFCS)[81] performed by clinical experts and confirmed by a Pediatric Neurologist. This approach reduces the risk of bias due to unblinding of the assessors to predictors (e.g., abnormal findings on brain imaging). Outcomes assessed at different occasions may introduce risk of bias, if the frequency of the assessment between participants varied. However, this is not the case for this study as participants were assessed at regular intervals (at 4, 8, 18 and 36 months of corrected age), and the diagnosis of CP was confirmed at or close to 36 months of corrected age. Additionally, there is low risk of selection bias with 100% of the birth cohort identified and 93% of eligible preterm survivors in this cohort having their neurodevelopmental assessments completed.

6.3 Predictors of CP

In the following section, the most pertinent predictors of CP that were identified in the current analysis will be discussed in the context of the relevant literature. While most of these predictors have physiologic plausibility for their relationship with CP, it should be emphasized that the current analysis was not designed to identify explanatory (i.e. causal) factors. Some of the studies cited below, however, aimed to assess explanatory factors, and hence, the comparability with the current study and predictors identified herein is limited.

6.3.1 Marital status

In this cohort, being married or in a common-law relationship was identified as an independent predictor of CP by logistic regression and was ranked the third on the variable importance plot of the prenatal model using RF. A recent Canadian study reported increased risk of mortality (18% vs. 11%, p = 0.009) and NDI (47% vs. 29%, p = 0.003) in preterm infants born to a single parent compared to those born to two-parent families (159). Almost one in five Canadian families are of single parent status, the majority (80-90%) of those being female single parents, with Nova Scotia having the highest rate of children living in lone parent families in Canada (160). There are two postulated pathways by which single parent status may adversely affect parental mental health and infant outcomes. Firstly, single mothers have high stress levels, which may disrupt the placental function of

maintaining maternal-fetal intermediary homeostasis, and adversely affect the developing fetal brain (161); this is supported by data from human and animal studies showing negative adverse effect of parental stress and uterine environment on the neuroplasticity of the developing brain, the parental bonding, and the subsequent psychosocial and behavioral development of the offspring (161). Secondly, single parents often have lower education, lower income, and less access to adequate prenatal care and they are more likely to face financial and social challenges compared to two-parent families (162, 163). Thirdly, parental stress may affect maternal perceptions, attitude and parent-child interaction/attachment, thus negatively impacting the social environment and neurosensory stimulation of the child and leads to behavioral and emotional adverse development (164, 165). Preterm infants may be at special risk of adverse behavioral and psychosocial outcomes related to parental stress compared to term infants, owing to their inherent biologic risk for brain injury and the vulnerability of their developing brains to environmental stressors. Maternal stress related to preterm birth, augmented by the lack of support from a partner, was associated with internalizing and externalizing behavior of their children at 3 years of corrected age (165).

6.3.2 Maternal age

Maternal age was identified as an independent predictor of CP in the RF model (see Figure 5.5.1 and Figure S3a in Appendix 2), but not in logistic regression. In the RF variable importance plot, maternal age ranked the first on the prenatal model and the 7th, out of 30 predictors, in the full prediction model of CP. Several studies and a recent systematic review reported an association between advanced maternal age (>35 years) and CP particularly in late preterm and term infants (166). However, other studies identified a nonlinear (U or J shaped) relation between maternal age at childbirth and development of CP in their offspring (167, 168). A recent report on 1391 children with CP from the Canadian CP Registry showed that 19% of those children were born to mothers aged 35 or older and 4% were born to mothers younger than 20 years (167). The Australian Early Development Census compared 107,666 aboriginal and non-aboriginal children and reported similar findings with a J-relation between maternal and neurodevelopment of the offspring at 5 years of age; being highest (40%, 95% CI 32-49) in children born to mothers

younger than 16 years of age and lowest (17%-18%) in children born to mothers 30-35 years of age, to increase again (reaching 17%-24%) for children born to mother older than 35 years of age (168). Both studies identified age-related socioeconomic or pregnancy risk factors at both extremes of maternal age. During logistic regression development, I examined the shape of the relationship between maternal age and the logit of CP and found it to be linear; therefore, I did not categorize the maternal age variable. The fact that RF identified maternal age as a predictor may be due to an interaction of maternal age with another variable that was strongly predictive of CP; such interaction was not considered in the main effects-only logistic regression model.

6.3.3 Antenatal steroids

Use of maternal antenatal steroids was identified as an independent predictor of CP in both the logistic regression and RF models. This finding is in agreement with a large body of evidence on the role of antenatal steroids in improving survival, reducing short-term morbidities (such as RDS, IVH, NEC), and improving long-term outcomes in preterm infants (169). A systematic review of clinical trials of single-course of antenatal steroids for preterm birth (before 34 weeks' gestation) showed a significant reduction in CP (7 studies, 146 of 1379 infants; RR 0.68, 95%CI 0.56-0.81), severe NDI (RR 0.79, 95%CI 0.73-0.80) and a significant improvement of intact survival (RR 1.19, 95%CI 1.06-1.33) (169). The exact mechanism is unknown, but animal studies showed that antenatal steroids resulted in maturation of the sympathoadrenal mechanisms involved in postnatal adaptation of preterm sheep, thus optimizing the metabolic, cardiac and respiratory responses to preterm birth (170).

6.3.4 Cystic white matter injury

Both RF and logistic regression showed that cystic white matter lesions (defined as cystic PVL and/or porencephaly) was the strongest predictor of CP in this cohort. The strong association between cystic white matter lesions and CP agrees with a systematic review and meta-analysis of 12 studies of CP in preterm children (79). The definition of cystic white matter lesions in the current analysis included both ischemic (cystic PVL) and hemorrhagic (porencephalic cyst as a consequence to parenchymal hemorrhage) lesions.

This is consistent with definitions of preterm white matter injury defined in literature based on abnormalities of brain imaging, with parenchymal hemorrhage, alone or in combination with cystic PVL, being the standard definition used by most studies (13-17, 21, 50-52, 155,156). Cystic PVL, also called "encephalopathy of prematurity", is a neuronal/axonal disease that affects the white matter, but may extend to include the thalamus, basal ganglia, brainstem and cerebellum (31). Cystic PVL is characterized by ischemia reperfusion injury, influx of inflammatory mediators, apoptosis and delayed neuronal maturation (31). With advances in perinatal care, the rates of cystic PVL have declined, but the rate of parenchymal hemorrhage remained unchanged at around 10 to 15% over the last decade. Therefore, this analysis and previous studies have commonly combined the devastating but rare cystic PVL with other cystic white matter lesions (such as porencephaly) into one exposure, particularly in studies with small sample size or with few cases of cystic PVL.

The diagnosis of brain lesions in this cohort was largely based on sequential bed-side cranial ultrasound done at regular intervals from birth until discharge or term age. At IWK, routine sequential cranial ultrasound is done on days 7, 14, and 42 after birth and at term-equivalent age, with more frequent scans performed if an abnormality is detected on routine screening. Sequential ultrasound up to term age provided high specificity and negative predictive value for the prediction of CP (sensitivity 76%, specificity 95%, PPV 48% and NPV 99%) (84). Although brain MRI at term-equivalent age has been used for prediction of CP in preterm children, (16, 87-89, 102, 109), it is not a standard practice at IWK; it is sometimes offered for high risk subgroup of preterm survivors, with established severe brain injury, hence it was not included as a predictor of CP in this study.

6.3.5 Other predictors of CP

Gestational age was not retained as a predictor for CP in any of the logistic regression models, in spite of the well-established inverse association between gestational age and CP. In the RF variable importance plot, Gestational age was ranked as 8th and 14th, among 30 predictors, in the combined prenatal-perinatal model and the full model, respectively. The lack of association between gestational age and CP was reported by 7 out of 9 studies included in a meta-analysis of CP in preterm children (79). The explanation for this finding

in this cohort is likely the presence of stronger downstream predictors of CP, such as cystic white matter lesions, that nullified the association between gestational age and CP. Another contributing factor may be the higher hospital mortality in extremely low gestational age infants, which may have removed infants from the sample that would otherwise be at high risk of CP.

The current study was limited to early clinical predictors that could be abstracted from the medical records prior to term age or hospital discharge. Therefore, it did not include some late predictors of CP reported in literature, such as assessment of general movements or other standardized neuromotor tests conducted over the first 6-12 months of post-term age (16,91,93,96, 97,102,109). The AIMS and the BSITD adapted screener are the only standardized neuromotor tests collected by the AC Allen Provincial PFUP database and are done routinely at 4-8 months of post-term age (105,140). I did not include those tests as predictors of CP in this study as they are performed beyond the postnatal time point (term age or hospital discharge) which was selected for early prediction of CP and early referrals to rehabilitation. Additionally, assessment of general movements was not included in this analysis as it was not performed in our center at the time the cohort was assembled.

6.4 Comparison of logistic regression and random forest

The main objectives of this study were to develop prediction models of CP using logistic regression and RF in a population-based cohort of very preterm children, to test their predictive performance individually and to examine whether a non-parametric model like RF would predict CP better than conventional logistic regression. In this section, I will discuss the performance of each prediction method individually and then compare their predictive performance to what is reported in the literature and to each other.

6.4.1 Logistic regression

Traditionally, logistic regression has been used in neonatal literature for the prediction of outcomes. The logistic regression model that was developed in this cohort provided good discrimination (AUC of 0.84) and accuracy (84%) in classifying children with and without CP. Compared to the thirteen prediction studies of CP in preterm children using logistic

regression listed in Table 3.1, only two studies (17,91) reported the discriminative ability of the developed model (AUC). Broitman et al. (2007) reported CP among 2103 surviving extreme low birth weight infants (<1000 grams) admitted to the National Neonatal Research Network (19 centers). The authors showed that isolated cranial ultrasound findings were poor predictors of CP compared to clinical models and that "All" model (clinical and ultrasound variables up to 36 weeks) improved CP prediction compared to "early" model (clinical and ultrasound variables at 28 days); AUC 0.78 vs AUC 0.72, respectively (P<0.01) (17). Ferrari et al. (2002) followed a small cohort of 84 preterm infants <37 weeks' (CP rate 49%) with abnormal cranial ultrasound findings (defined as cystic or non-cystic white matter lesions or grade 3 IVH) and examined late clinical variables (general movement assessment and standardized neuromotor exams over the first 5 months) as predictors of CP at 2-3 years of age (93). The study showed that consistent cramped synchronized movements to accurately predict CP and reported a significant difference in the discrimination between general movement and ultrasound abnormality as predictors of CP; AUC 0.97 and AUC 0.88, respectively (P = 0.001).

The current study reported a higher AUC compared to that by Broitman et al (AUC 0.84 vs 0.78). Both studies used clinical predictors up to hospital discharge or 36 weeks, including ultrasound abnormalities. However, Broitman et al included smaller extreme preterm infants at higher risk of CP (CP rate 16%) and reported CP as part of composite outcome of NDI in a multicentre cohort with 23% loss to follow up rate, compared to this analysis of population-based cohort of bigger preterm infants with low attrition of 7%.

Although the AUC reported by this analysis is inferior to that by Ferrari et al, both studies are not comparable: (i) Ferrari et al included a small cohort of higher gestational age infants compared to the population-based cohort of very preterm infants (<31 weeks) in this analysis, (ii) the study had high risk of selection bias (CP rate 49%); including only those infants with ultrasound abnormalities, as a selective small high-risk subgroup of their preterm cohort (iii) and finally, the authors assessed mainly late predictors compared to early clinical predictors in this analysis.

The added benefit of including late predictors, particularly general movement assessment and standardized neuromotor tests, to brain imaging in prediction of CP in the preterm population has been reported elsewhere. Morgan et al. (2019) conducted a retrospective case control study on 441 high risk preterm and term infants and combined all three predictors (neuroimaging, general movements assessment and Hammersmith Infant Neurological Examination scores at 3 months of post-term age) to develop a pooled early prediction model of CP (111). The authors reported a pooled prediction with AUC of 0.99, which was higher than that for any individual predictor with excellent classification accuracy (98.7%) and diagnostic properties (sensitivity 97.9; specificity 99; PPV 98.6; NPV 98.8) (111). Including late variables in future prediction studies of CP may improve the discriminative abilities of CP prediction models in preterm children. As of 2020, many follow-up programs in Canada, including the Nova Scotia PFUP, routinely assess the general movements in preterm infants, so it could be included in future prediction studies of CP in Nova Scotia.

The classification accuracy of the full logistic regression model in this analysis is comparable to three small studies reporting accuracy of CP classification of 74%, 80% and 94% (16, 99, 106). All these studies were limited by small sample size, low absolute number of CP cases (< 10 per study), and using late clinical predictors (general movements assessment, term neuroimaging/behavioral assessment, combined motor tests) assessed at 3-12 months of post-term age.

The diagnostic properties of the full logistic regression in this study has a high NPV of 97% and low false positive rate (15%) which makes it useful for clinical use. The PPV of the full model was low (40%) which is consistent with the findings from eight prediction studies of CP using logistic regression and reporting low PPV ranging from 19%-60% (16, 86, 89, 91, 99, 104, 106). However, with the exception of AUC, caution with the interpretation of the diagnostic properties of the developed prediction models is advised as they depend on the selected cutoff and would result in different values for the same sample if a different cutoff was selected.

6.4.2 Machine learning

Over the last two decades, health researchers have increasingly used machine learning to develop diagnostic or prognostic models using clinical variables. In neonatal research, two small studies used multidimensional data to develop machine learning prediction of CP in children born preterm. One study by Hope et al (2008) used RF for prediction of CP (85), whereas the other study used support vector machines (101). The study that used RF was a single centre retrospective Canadian study, conducted at the IWK Health Centre. The authors predicted CP from quantitative texture measures of early cranial ultrasound scans performed within the first seven postnatal days. The study was limited by the small sample size (37 CP cases and 48 controls) and the case-control design (85). The authors did not report AUC or the diagnostic properties of their CP prediction model, but they reported the classification accuracy which is lower compared to the full model in the current analysis (72% vs 91%, respectively) (83). Stahl et al. prospectively followed 82 infants at 10-18 weeks post-term age and predicted CP from the infants' movement patterns; they reported that a combination of three motion image variables was the most accurate in predicting CP (85% sensitivity, 88% specificity, AUC 0.88 (95% CI 0.77–1.00) (101).

The main difference between this analysis and those studies is their use of machine learning for CP prediction in the context of multidimensional data, such as video analysis of movements or texture analysis of ultrasound images, in contrast with the current analysis testing RF predictive validity relying only on clinical predictors.

The developed prediction model of CP using RF in the present study has many advantages over other prediction studies of CP using machine learning methods: (i) being the first population-based study that used RF to predict CP in preterm children and to compare the predictive performance to conventional logistic regression in this context; (ii) being the first study to test RF based only on clinical predictors that can be abstracted from patient records compared to the multidimensional data used by other studies; (iii) the low risk of selection bias compared to the other studies that included only a selected high risk subgroup of preterm infants with CP rates being much higher than the general population; and (iv)

using early predictors, prior to hospital discharge or before term age, whereas Stahl et al. used relatively late predictors occurring between 10 and 18 weeks of corrected age (101).

6.4.3 Comparing the predictive performance of RF and logistic regression models of CP

In the present study, both RF and logistic regression yield comparable prediction of CP using clinical predictors with regards to discrimination and classification. Both methods resulted in good discrimination with similar AUC (0.83 vs. 0.84) and slightly better classification favoring RF (91% vs 84%). For binary outcomes with low dimensional data (the number of covariates is small relative to the sample size), logistic regression is the conventional statistical method used for prediction, particularly when researchers are interested in explanation (i.e. estimating the causal association between a risk factor and an outcome) in addition to prediction. The existing literature comparing different machine learning methods (RF, neuronal network, support vector machines and gradient boosted decision trees) over conventional regression (logistic or Cox-regression) for prediction of clinical outcomes reported conflicting results from superior (131-135) to similar or even inferior discrimination (136-138).

The lack of additional benefit of RF over logistic regression in the current study has been recently reported in a handful of studies, particularly when clinical variables were used (111, 136-138, 171). Pua et al. followed a cohort of 4026 adult patients and compared seven different machine learning methods to logistic regression for prediction of walking limitation after total knee arthroplasty using demographic and clinical variables similar to this analysis. The authors reported similar discrimination of the ordinal logistic regression and RF (AUC 0.75 vs. 0.74) and suggested that this could be expected to occur when the predictors act additively (i.e. there are no interactions between predictors) or when non-linearity is not substantial enough for machine learning to be of additional benefit (171).

The conflicting reports in literature, from superior to inferior discrimination of RF compared to conventional regression, suggest that none of the methods is superior and that

the predictive validity may rely on the settings or the datasets and may not be constant across different studies. A recent systematic review of 71 studies published between January 2016 and August 2017 reported similar predictive accuracy of machine learning and logistic regression in the group of studies with low risk of bias, but superior predictive accuracy of machine learning in the group of studies at high risk of bias (172). Almost half (137/282) of the included studies in the review were at high risk of bias that was attributed to poor methodology or poor reporting of variable selection procedures, the number of predictors, checking for linearity and interactions between continuous predictors, dealing with class imbalance of the outcome, or validation (172).

Recently, Couronne et al. designed a benchmarking experiment, using 243 real-life datasets to compare the predictive ability between logistic regression and RF using the standard RF variant with the default tuning parameters as implemented in the widely used R package *randomForests* for pragmatic comparisons (173). The authors showed that RF was superior to logistic regression in approximately 69% of the datasets for AUC and accuracy, but the difference between both methods was small. The authors also observed that certain characteristics of the dataset such as the sample size and the number of predictors were associated with superior accuracy of RF over conventional regression when the number of predictors was ≥ 5 or the ratio of predictors to sample size was > 0.1 (173).

Apart from discrimination (AUC), caution when interpreting the diagnostic properties of the developed prediction models of CP in the current study is advised. For logistic regression models, I selected the cutoff that maximized the sensitivity, specificity and correct classification. However, different cutoffs would result in different diagnostic properties. For RF, I used the confusion matrix to obtain the diagnostic properties of the developed RF models. However, this has to be interpreted with caution in view of the correction for the class imbalance in the dataset. As the classification algorithms make assumptions that the test data are drawn from the same class distribution of the training data, RF will favor the majority class in presence of class imbalance. This makes it challenging to create appropriate testing and training data sets, unless correction of the

imbalance is performed. In this analysis, the minority class (CP) was up-weighed to create even groups, before splitting of the study sample.

6.5 Strengths and Limitations

This study provides insight on a novel prediction method of CP using RF applied to a population-based cohort of very preterm children and it compared its predictive performance to the traditionally used logistic regression. With increased survival of extreme preterm infants at highest risk for CP, research that focuses on developing accurate prediction models of CP in those infants is very valuable for patients, their families and healthcare providers.

This study has several strengths. First, a major strength is the use of a population-based cohort of very preterm children. Second, the attrition was low: 93% of the eligible cohort had completed the outcome assessment improving the validity of the results, particularly with the high loss to follow-up (20-50%) reported by previous population-based studies of preterm children (16,17,88,91,98,99). Third, the ascertainment of the primary outcome in this study was based on multiple neurological assessments by experienced care providers and confirmed by expert neurologists with standardized grading of CP severity based on a validated classification system (GMFCS). The primary outcome was ascertained close to 36 months corrected age, which increases the robustness of the diagnosis and improves the diagnostic accuracy of CP, particularly for the mild ambulatory subtype. Finally, the analysis used a split sample for internal validation of the developed prediction models.

Compared to the published studies of CP prediction in preterm children [Table 3.1], this study reported the measures of performance of the developed models (discrimination, calibration and correct classification) in addition to their diagnostic properties. Importantly, none of the studies in Table 3.1 performed internal validation to test the performance of the developed prediction models in a similar but independent sample from the same population.

The development and reporting of CP prediction using RF and logistic regression in this cohort could be considered superior compared to some of the published studies despite the retrospective design. This could be attributed to multiple factors: (i) the low risk of selection bias because it was a large population-based cohort with excellent follow up rate, compared to the majority of reported studies derived from small cohorts; (ii) inclusion of very preterm infants (< 31 weeks) who are at highest risk of CP compared to bigger preterm infants included in other studies (86,90,93,95,98-100), particularly the largest two studies (429 infants <33 weeks' gestation and 903 infants < 37 weeks' gestation) (86,95); (iii) the predictors identified by this study have established physiological plausibility in the development of CP and the fact that both RF and logistic regression yield similar predictors at each time point, reflects the robustness of these predictors and the validity of the developed models.

This study used early clinical predictors of CP, starting from the prenatal period, through the perinatal and postnatal periods and up to hospital discharge or the expected date of delivery. This in contrast with the majority of prediction studies of CP in preterm children relying on late predictors: Thirteen studies in Appendix 1 used relatively late predictors (4 studies at term age, 7 studies at 3-4 months post term age and 2 studies over the first 6-12 months). The clinical implications and objectives of CP prediction at different time points vary considerably, both for families and caregivers. Early prediction at birth or during the first postnatal days are crucial to guide critical discussions around provision or withdrawal of life support or intensive care for extremely preterm or critically ill infants. The full model in this study provided prediction at or near-term age to enable early referral to targeted interventions and individualized care planning that impact both patient outcomes and health services utilization.

This study has some limitations that need to be acknowledged. The retrospective nature of the data excluded potential predictors that were not included in the database (e.g., general movement assessment), which may have affected the selected predictors and the predictive performance. Some known predictors of CP (such as race/ethnicity) were not included in the analysis, because of inconsistency in data collection that would have impacted the

internal validity of the study. The fact that the majority of the population in Nova Scotia are Caucasians limits the generalizability of the findings from this analysis to other populations with different ethnic groups or race-related determinants of health.

Missing data is a known limitation, but the proportion of missing values in this study was small. Only SES had more than 5% missing values (11%). If SES is missing not at random, e.g. if a parent's decision to report SES information depends on their SES, estimates from the model's will be biased, which in turn would adversely affect the model's predictive performance in new data. Additionally, some variables known to be associated with mortality and CP in preterm infants were not included in the study, because of a large number of missing values (e.g. neonatal severity of illness scores, physiologic definition of BPD). However, the model contained several predictors that are closely related to the ones omitted, so that the negative effect on model performance is likely negligible.

There is a possibility of misclassification with some children that were classified as having mild or suspected CP would have been classified with motor delay without CP, if followed beyond 3 years of corrected age. A Canadian cohort of preterm children born before 29 weeks' gestation with suspected CP at 18 months of corrected age showed that their developmental trajectory at 3 years of age was midway between those with CP and normal children (174). Conversely, preterm children with mild CP may only exhibit clinical signs after 36 months of corrected age. Although this remains a possibility, the identification of CP in this cohort was based on multiple assessments by experienced clinicians, using standardized tests, and those with suspected CP were referred to Pediatric neurologists for confirmation of the diagnosis. Only those with normal neuromotor exams on multiple assessments up to 18-24 months of corrected age receive standardized neurodevelopmental assessment with BSITD at 3 years. If a subtle abnormality in motor development is detected or a parental concern is raised at that time, a detailed neurologic exam would be repeated to confirm the findings. Paneth et al (2006) argued from a public health perspective that those children with subtle or subclinical CP should not be counted as CP cases as they lack the social, familial and medical burden of a typical child with CP (8).

6.6 Implications for Clinical Practice and Future Research Directions

In this cohort, the full RF model provides similar discrimination and accuracy compared to logistic regression for prediction of CP in very preterm children. If external validation of the developed RF model (very preterm birth cohort from PEI) shows superior predictive performance, then this model may be used as adjunct to improve the identification of preterm children at risk of CP to target early intervention and to make efficient use of the limited health care resources. Regardless of the prediction method used, the poor discrimination of the prenatal and perinatal models do not allow for their use in prediction or for counselling parents at the respective timepoints. Additionally, clinicians should be aware of the false positive and false negative results of these prediction models when counselling families of preterm infants. Therefore, these prediction models should not be used in isolation, but as adjunct to other clinical parameters to aid in the diagnosis of CP.

Parents are often interested in individualized timely prediction of CP or NDI of their preterm child. Caregivers are often asked about this prognostic information at various times: prior to preterm birth, within the first few days following NICU admission, and at hospital discharge. The prediction models developed in the present study may be used to predict an infant's probability of developing CP given his/her set of predictors. The predicted probability of developing CP can be easily transformed into calculated risk-based scores and algorithms. The clinical calculators derived from such prediction models have been widely used in neonatal practice to aid for counselling families or when critical decisions are discussed and can be made available for use by clinicians at hand.

Future research should continue to explore the role of machine learning methods in prediction of CP and other clinical outcomes using existing datasets. The Canadian Neonatal Network database collects maternal and infant data for an average of 2500 preterm infants < 33 weeks or VLBW infants < 1500 g admitted to all 31 tertiary level Canadian neonatal units every year. The long-term follow-up data of a subgroup of survivors who were born before 29 weeks' gestation (around 1000 infants/year) is linkable through the Canadian Perinatal Follow up Network collecting data from 28 regional follow

up programs across Canada (175). These programs collect and share parental demographic and socioeconomic data, children's general health and growth data, neurodevelopmental outcomes and health services utilization at 18-24 months corrected age for research purposes (175). Accurate risk prediction of preterm survivors would help to redirect resources toward those who are most likely to benefit, including rehabilitation, family resources, and social support. Tools should be developed to implement the use of machine learning prediction models at the bedside.

References

1. Statistics Canada. Preterm live births in Canada 2000-2013. 2016. Available from https//www.statcan.gc.ca/pub/82-625-x/2016001/article/14675-eng.htm Accessed May 4, 2021.

2. Lui K, Lee SK, Kusuda S, Adams M, Vento M, Reichman B, et al. Trends in outcomes for neonates born very preterm and very low birth weight in 11 high-income countries. J Pediatr. 2019; 215:32-40.e14.

3. Behrman Re BAS. Preterm birth: causes, consequences, and prevention 2007. Available from: http://www.nap.edu/catalog/11622/preterm-birth-causes-consequences-and-prevention. Accessed May 4, 2021.

4. Hee Chung E, Chou J, Brown KA. Neurodevelopmental outcomes of preterm infants: a recent literature review. Transl Pediatr. 2020;9(Suppl 1):S3-S8.

5. Perrott S, Dodds L, Vincer M. A population-based study of prognostic factors related to major disability in very preterm survivors. J Perinatol**.** 2003(2):111.

6. Synnes A, Luu TM, Moddemann D, Church P, Lee D, Vincer M, et al. Determinants of developmental outcomes in a very preterm Canadian cohort. Arch Dis Child Fetal Neonatal Ed. 2017;102(3):F235-F4.

7. Hollanders JJ, Schaëfer N, van der Pal SM, Oosterlaan J, Rotteveel J, Finken MJJ, et al. Long-term neurodevelopmental and functional outcomes of infants born very preterm and/or with a very low birth weight. Neonatology. 2019;115(4):310-9.

8. Paneth N, Hong T, Korzeniewski S. The descriptive epidemiology of cerebral palsy. Clin Perinatol. 2006(2):251.

9. Christensen D, Van Naarden Braun K, Doernberg NS, Maenner MJ, Arneson CL, Durkin MS, et al. Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning. Dev Med Child Neurol. 2014;56(1):59-65.

10. Johnston KM, Gooch K, Korol E, Vo P, Eyawo O, Bradt P, et al. The economic burden of prematurity in Canada. BMC Pediatr. 2014;14:93.

11. Hadders-Algra M. Early diagnosis and early intervention in cerebral palsy. Front Neurol. 2014;5:185.

12. Philippi H, Karch D, Kang KS, Wochner K, Pietz J, Dickhaus H, et al. Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy. Dev Med Child Neurol. 2014;56(10):960-7.

13. Maitre NL, Marshall DD, Price WA, Slaughter JC, O'Shea TM, Maxfield C, et al. Neurodevelopmental outcome of infants with unilateral or bilateral periventricular hemorrhagic infarction. Pediatrics. 2009;124(6):e1153.

14. de Vries LS, van Haastert IC, Benders MJ, Groenendaal F. Myth: cerebral palsy cannot be predicted by neonatal brain imaging. Semin Fetal Neonatal Med**.** 2011;16(5):279.

15. Cioni G, Fazzi B, Coluccini M, Bartalena L, Boldrini A, van Hof-van Duin J. Cerebral visual impairment in preterm infants with periventricular leukomalacia. Pediatr Neurol. 1997;17(4):331.

16. Constantinou JC, Adamson-Macedo EN, Mirmiran M, Fleisher BE. Movement, imaging and neurobehavioral assessment as predictors of cerebral palsy in preterm infants. J Perinatol. 2007;27(4):225-9.

17. Broitman E, Ambalavanan N, Higgins RD, Vohr BR, Das A, Bhaskar B, et al. Clinical data predict neurodevelopmental outcome better than head ultrasound in extremely low birth weight infants. J Pediatr. 2007;151(5):500-5, 5.e1-2.

18. Tyson JE, Parikh NA, Langer J, Green C, Higgins RD, Network NIoCHaHDNR. Intensive care for extreme prematurity - moving beyond gestational age. N Engl J Med. 2008;358(16):1672-81.

19. Moore T, Hennessy EM, Myles J, Johnson SJ, Draper ES, Costeloe KL, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. BMJ. 2012;345:e7961.

20. Leversen KT, Sommerfelt K, Rønnestad A, Kaaresen PI, Farstad T, Skranes J, et al. Prediction of neurodevelopmental and sensory outcome at 5 years in Norwegian children born extremely preterm. Pediatrics. 2011;127(3):e630-8.

21. Beaino G, Khoshnood B, Kaminski M, Pierrat V, Marret S, Matis J, et al. Predictors of cerebral palsy in very preterm infants: the EPIPAGE prospective population-based cohort study. Dev Med Child Neurol. 2010;52(6):e119-25.

22. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and Regression Trees. Boca Raton: CRC Press,1984.

23. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

24. Mandaleson A, Lee Y, Kerr C, Graham HK. Classifying cerebral palsy: are we nearly there? J Pediatr Orthop. 2015;35(2):162-6.

25. Bax M, Goldstein M, Rosenbaum P, Leviton A, Paneth N, Dan B, et al. Proposed definition and classification of cerebral palsy, April 2005. Dev Med Child Neurol. 2005;47(8):571-6.

26. Odding E, Roebroeck ME, Stam HJ. The epidemiology of cerebral palsy: incidence, impairments and risk factors. Disabil Rehabil. 2006;28(4):183-91.

27. Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernandez J, et al. Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. JAMA Pediatr. 2017;171(9):897-907.

28. Platt MJ, Cans C, Johnson A, Surman G, Topp M, Torrioli MG, et al. Trends in cerebral palsy among infants of very low birthweight (<1500 g) or born prematurely (<32 weeks) in 16 European centres: a database study. Lancet. 2007;369(9555):43-50.

29. Vincer MJ, Allen AC, Allen VM, Baskett TF, O'Connell CM. Trends in the prevalence of cerebral palsy among very preterm infants (<31 weeks' gestational age). Paediatr Child Health. 2014(- 4):185.

30. O'Shea TM, Dammann O. Antecedents of cerebral palsy in very low-birth weight infants. Clin Perinatol. 2000;27(2):285-302.

31. Babcock MA, Kostova FV, Ferriero DM, Johnston MV, Brunstrom JE, Hagberg H, et al. Injury to the preterm brain and cerebral palsy: clinical aspects, molecular mechanisms, unanswered questions, and future research directions. J Child Neurol. 2009;24(9):1064-84.

32. Coq JO, Delcour M, Massicotte VS, Baud O, Barbe MF. Prenatal ischemia deteriorates white matter, brain organization, and function: implications for prematurity and cerebral palsy. Dev Med Child Neurol. 2016;58 Suppl 4:7-11.

33. Schneider J, Miller SP. Preterm brain Injury: White matter injury. Handb Clin Neurol. 2019;162:155-72.

34. Duerden EG, Taylor MJ, Miller SP. Brain development in infants born preterm: looking beyond injury. Semin Pediatr Neurol. 2013;20(2):65-74.

35. Shatrov JG, Birch SCM, Lam LT, Quinlivan JA, McIntyre S, Mendz GL. Chorioamnionitis and cerebral palsy: a meta-analysis. Obstet Gynecol. 2010;116(2 Pt 1):387-92.

36. Pappas A, Kendrick DE, Shankaran S, Stoll BJ, Bell EF, Laptook AR, et al. Chorioamnionitis and early childhood outcomes among extremely low-gestational-age neonates. JAMA Pediatr. 2014;168(2):137-47.

37. Kaukola T, Herva R, Perhomaa M, Paakko E, Kingsmore S, Vainionpaa L, et al. Population cohort associating chorioamnionitis, cord inflammatory cytokines and neurologic outcome in very preterm, extremely low birth weight infants. Pediatr Res. 2006;59(3):478.

38. Gray PH, Jones P, O'Callaghan MJ. Maternal antecedents for cerebral palsy in extremely preterm babies: a case-control study. Dev Med Child Neurol. 2001;43(9):580-5.

39. Grobman WA, Lai Y, Rouse DJ, Spong CY, Varner MW, Mercer BM, et al. The association of cerebral palsy and death with small-for-gestational-age birthweight in preterm neonates by individualized and population-based percentiles. Am J Obstet Gynecol. 2013;209(4):340 e1-5.

40. Freire G, Shevell M, Oskoui M. Cerebral palsy: phenotypes and risk factors in term singletons born small for gestational age. Eur J Paediatr Neurol**.** 2015;19(2):218.

41. O'Callaghan ME, MacLennan AH, Gibson CS, McMichael GL, Haan EA, Broadbent JL, et al. Epidemiologic associations with cerebral palsy. Obstetrics and Gynecology. 2011;118(3):576.

42. Leppanen M, Ekholm E, Palo P, Maunu J, Munck P, Parkkola R, et al. Abnormal antenatal Doppler velocimetry and cognitive outcome in very-low-birth-weight infants at 2 years of age. Ultrasound in Obstet Gynecol. 2010;36(2):178.

43. Skiöld B, Alexandrou G, Padilla N, Blennow M, Vollmer B, Adén U. Sex differences in outcome and associations with neonatal brain morphology in extremely preterm children. J Pediatr. 2014;164(5):1012-8.

44. Peacock JL, Marston L, Marlow N, Calvert SA, Greenough A. Neonatal and infant outcome in boys and girls born very prematurely. Pediatric research. 2012;71(3):305.

45. McMillan EM, King GM, Adamson IY. Sex hormones influence growth and surfactant production in fetal lung explants. Exp Lung Res. 1989;15:167.

46. Counsell SJ, Boardman JP. Differential brain growth in the infant born preterm: current knowledge and future developments from brain imaging. Semin Fetal Neonatal Med. 2005;10(5):403-10.

47. Hagberg H MCJB. Role of cytokines in preterm labour and brain injury. BJOG. 2005;112:16.

48. Hüppi PS, Warfield S, Kikinis R, Barnes PD, Zientara GP, Jolesz FA, et al. Quantitative magnetic resonance imaging of brain development in premature and mature newborns. Ann Neurol. 1998;43(2):224-35.

49. Zhang S, Bingbing Li, Zhang X, Zhu C, Wang X. Birth Asphyxia Is Associated with Increased Risk of Cerebral Palsy: A Meta-Analysis. Front Neurol. 2020 Jul 16;11:704.

50. Nelson KB. Intrapartum asphyxia: a rare cause of cerebral palsy. J Pediatr. 1988;112 (4): 515-519.

51. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. J Pediatr. 1978;92(4):529-34.

52. Gotardo JW, Volkmer NFV, Stangler GP, Dornelles AD, Bohrer BBA, Carvalho CG. Impact of peri-intraventricular haemorrhage and periventricular leukomalacia in the neurodevelopment of preterms: a systematic review and meta-analysis. PLoS One. 2019 Oct 10;14(10):e0223427.

53. Resic B, Tomasovic M, Kuzmanic-Samija R, Lozic M, Resic J, Solak M. Neurodevelopmental outcome in children with periventricular leukomalacia. Coll Antropol. 2008;32 Suppl 1:143.

54. Shang Q, Ma CY, Lv N, Lv ZL, Yan YB, Wu ZR, et al. Clinical study of cerebral palsy in 408 children with periventricular leukomalacia. Exp Ther Med. 2015;9(4):1336.

55. Zaghloul N, Ahmed M. Pathophysiology of periventricular leukomalacia: What we learned from animal models. Neural Regen Res. 2017;12(11):1795-6.

56. Cai S, Thompson DK, Anderson PJ, Yang JY. Short- and Long-Term Neurodevelopmental Outcomes of Very Preterm Infants with Neonatal Sepsis: A Systematic Review and Meta-Analysis. Children (Basel). 2019 Dec 1;6(12):131.

57. Schlapbach LJ, Aebischer M, Adams M, Natalucci G, Bonhoeffer J, Latzin P, et al. Impact of sepsis on neurodevelopmental outcome in a Swiss National Cohort of extremely premature infants. Pediatrics. 2011;128(2):e348-57.

58. Wang LW, Lin YC, Wang ST, Yeh TF, Huang CC. Hypoxic/ischemic and infectious events have cumulative effects on the risk of cerebral palsy in very-low-birth-weight preterm infants. Neonatology. 2014;106:209.

59. Hansen ML, Jensen IV, Gregersen R, Juhl SM, Greisen G. Behavioural and neurodevelopmental impairment at school age following necrotising enterocolitis in the newborn period. PLoS One. 2019;14(4): e0215220.

60. Pike K, Brocklehurst P, Jones D, Kenyon S, Salt A, Taylor D, Marlow N. Outcomes at 7 years for babies who developed neonatal necrotising enterocolitis: the ORACLE Children Study. Arch Dis Child Fetal Neonatal Ed. 2012;5:F318.

61. Gough A, Linden MA, Spence D, Halliday HL, Patterson CC, McGarvey L. Executive functioning deficits in young adult survivors of bronchopulmonary dysplasia. Disabil Rehabil. 2015;37(21):1940-5.

62. Van Marter LJ, Kuban KC, Allred E, Bose C, Dammann O, O'Shea M, et al. Does bronchopulmonary dysplasia contribute to the occurrence of cerebral palsy among infants born before 28 weeks of gestation? Arch Dis Child Fetal Neonatal Ed. 2011;96(1):F20.

63. Doyle LW, Halliday HL, Ehrenkranz RA, Davis PG, Sinclair JC. An update on the impact of postnatal systemic corticosteroids on mortality and cerebral palsy in preterm infants: effect modification by risk of bronchopulmonary dysplasia. J Pediatr. 2014;165(6):1258-60.

64. Doyle LW, Ehrenkranz RA, Halliday HL. Early (< 8 days) postnatal corticosteroids for preventing chronic lung disease in preterm infants. Cochrane Database Syst Rev. 2014(5):CD001146.

65. Lim G, Lee BS, Choi YS, Park HW, Chung ML, Choi HJ, et al. Delayed dexamethasone therapy and neurodevelopmental outcomes in preterm infants with bronchopulmonary dysplasia. Pediatr Neonatol. 2015;56(4):261-7.

66. Doyle LW, Cheong JL, Ehrenkranz RA, Halliday HL. Late (> 7 days) systemic postnatal corticosteroids for prevention of bronchopulmonary dysplasia in preterm infants. Cochrane Database Syst Rev. 2017;10:CD001145.

67. He J, Varma A, Weissfeld LA, Devaskar SU. Postnatal glucocorticoid exposure alters the adult phenotype. Am J Physiol Reg Integ Comp Physiol. 2004;287(1):R198-208.

68. Murphy BP, Inder TE, Huppi PS, Warfield S, Zientara GP, Kikinis R, et al. Impaired cerebral cortical gray matter growth after treatment with dexamethasone for neonatal chronic lung disease. Pediatrics. 2001;107(2):217.

69. Sood BG, Madan A, Saha S, Schendel D, Thorsen P, Skogstrand K, et al. Perinatal systemic inflammatory response syndrome and retinopathy of prematurity. Pediatr Res. 2010;67(4):394-400.

70. Hong HK, Lee HJ, Ko JH, Park JH, Park JY, Choi CW, et al. Neonatal systemic inflammation in rats alters retinal vessel development and simulates pathologic features of retinopathy of prematurity. J Neuroinflammation. 2014;11:87.

71. Goyen TA, Todd DA, Veddovi M, Wright AL, Flaherty M, Kennedy J. Eye-hand coordination skills in very preterm infants <29 weeks gestation at 3 years: Effects of preterm birth and retinopathy of prematurity. Early Hum Dev. 2006;82(11):739-45.

72. Goyen TA, Lui K. Developmental coordination disorder in "apparently normal" schoolchildren born extremely preterm. Arch Dis Child. 2009;94(4):298-302.

73. Leversen KT, Sommerfelt K, Rønnestad A, Kaaresen PI, Farstad T, Skranes J, et al. Predicting neurosensory disabilities at two years of age in a national cohort of extremely premature infants. Early Hum Dev. 2010;86(9):581-6.

74. Basu AP, Pearse J, Kelly S, Wisher V, Kisler J. Early intervention to improve hand function in hemiplegic cerebral palsy. Front Neurol. 2014;5:281.

75. Basu AP, Clowry G. Improving outcomes in cerebral palsy with early intervention: new translational approaches. Front Neurol. 2015;6:24.

76. Spittle AJ, Anderson PJ, Lee KJ, Ferretti C, Eeles A, Orton J, et al. Preventive care at home for very preterm infants improves infant and caregiver outcomes at 2 years. Pediatrics. 2010;126(1):e171-8.

77. Colditz P, Sanders MR, Boyd R, Pritchard M, Gray P, O'Callaghan MJ, et al. Prem Baby Triple P: a randomised controlled trial of enhanced parenting capacity to improve developmental outcomes in preterm infants. BMC Pediatr. 2015;15:15.

78. Mei C, Reilly S, Reddihough D, Mensah F, Green J, Pennington L, et al. Activities and participation of children with cerebral palsy: parent perspectives. Disabil Rehabil. 2015;37(23):2164-73.

79. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Prognostic factors for cerebral palsy and motor impairment in children born very preterm or very low birthweight: a systematic review. Dev Med Child Neurol. 2016;58(6):554-69.

80. Zhang J. Multivariate Analysis and Machine Learning in Cerebral Palsy Research. Front Neurol. 2017;8:715.

81. Bosanquet M, Copeland L, Ware R, Boyd R. A systematic review of tests to predict cerebral palsy in young children. Dev Med Child Neurol. 2013;55(5):418-26.

82. Wikström S, Pupp IH, Rosén I, Norman E, Fellman V, Ley D, et al. Early single-channel aEEG/EEG predicts outcome in very preterm infants. Acta Paediatr. 2012;101(7):719-26.

83. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol. 1997;39(4):214-23.

84. Schwindt E, Thaller C, Czaba-Hnizdo C, Giordano V, Olischar M, Waldhoer T, et al. Being born small for gestational age influences amplitude-integrated electroencephalography and later outcome in preterm infants. Neonatology. 2015;108(2):81-7.

85. Hope TA, Gregson PH, Linney NC, Schmidt MH, Abdolell M. Selecting and assessing quantitative early ultrasound texture measures for their association with cerebral palsy. IEEE Trans Med Imaging. 2008;27(2):228-36.

86. De Vries LS, Van Haastert IL, Rademaker KJ, Koopman C, Groenendaal F. Ultrasound abnormalities preceding cerebral palsy in high-risk preterm infants. J Pediatr. 2004;144(6):815-20.

87. Hagberg G, Olow I. The changing panorama of cerebral palsy in Sweden 1954-1970. II. Analysis of the various syndromes. Acta Paediatr Scand. 1975;64(2):193-200.

88. Lacey JL, Rudge S, Rieger I, Osborn DA. Assessment of neurological status in preterm infants in neonatal intensive care and prediction of cerebral palsy. Aust J Physiother. 2004;50(3):137-44.

89. Woodward LJ, Anderson PJ, Austin NC, Howard K, Inder TE. Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. NEJM. 2006;355(7):685.

90. Nanba Y, Matsui K, Aida N, Sato Y, Toyoshima K, Kawataki M, et al. Magnetic resonance imaging regional T1 abnormalities at term accurately predict motor outcome in preterm infants. Pediatrics. 2007;120(1):e10-9.

91. Mirmiran M, Barnes PD, Keller K, Constantinou JC, Fleisher BE, Hintz SR, et al. Neonatal brain magnetic resonance imaging before discharge is better than serial cranial ultrasound in predicting cerebral palsy in very low birth weight preterm infants. Pediatrics. 2004;114(4):992-8.

92. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl. 2007;109:8-14.

93. Ferrari F, Cioni G, Einspieler C, Roversi MF, Bos AF, Paolicelli PB, et al. Cramped synchronized general movements in preterm infants as an early marker for cerebral palsy. Arch Pediatr Adolesc Med. 2002;156(5):460-7.

94. Ellenberg JH, Nelson KB. Early recognition of infants at high risk for cerebral palsy: examination at age four months. Dev Med Child Neurol. 1981;23(6):705-16.

95. Romeo DM, Guzzetta A, Scoto M, Cioni M, Patusi P, Mazzone D, et al. Early neurologic assessment in preterm-infants: integration of traditional neurologic

examination and observation of general movements. Eur J Paediatr Neurol. 2008;12(3):183-9.

96. Haataja L, Cowan F, Mercuri E, Bassi L, Guzzetta A, Dubowitz L. Application of a scorable neurologic examination in healthy term infants aged 3 to 8 months. J Pediatr. 2003;143(4):546.

97. Voigt RG, Brown FR, Fraley JK, Llorente AM, Rozelle J, Turcich M, et al. Concurrent and predictive validity of the cognitive adaptive test/clinical linguistic and auditory milestone scale (CAT/CLAMS) and the Mental Developmental Index of the Bayley Scales of Infant Development. Clin Pediatr (Phila). 2003;42(5):427-32.

98. Øberg GK, Jacobsen BK, Jørgensen L. Predictive value of general movement assessment for cerebral palsy in routine clinical practice. Phys Ther. 2015;95(11):1489-95.

99. De Bock F, Will H, Behrenbeck U, Jarczok MN, Hadders-Algra M, Philippi H. Predictive value of General Movement Assessment for preterm infants' development at 2 years - implementation in clinical routine in a non-academic setting. Res Dev Disabil. 2017;62:69-80.

100. Adde L, Helbostad JL, Jensenius AR, Taraldsen G, Grunewaldt KH, Støen R. Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study. Dev Med Child Neurol. 2010;52(8):773-8.

101. Stahl A, Schellewald C, Stavdahl Ø, Aamo OM, Adde L, Kirkerød H. An optical flow-based method to predict infantile cerebral palsy. IEEE Trans Neural Syst Rehabil Eng. 2012;20(4):605-14.

102. Korner A. Neurobehavioural assessment of the preterm infant. In: Thom V, editor.: Child Development; 2000.

103. Amiel-Tison C. Neuromotor status. Follow-up management of the high-risk infant. Boston: Little, Brown and Company,1987. p. 115-26.

104. Skiöld B, Eriksson C, Eliasson AC, Adén U, Vollmer B. General movements and magnetic resonance imaging in the prediction of neuromotor outcome in children born extremely preterm. Early Hum Dev. 2013;89(7):467-72.

105. Europe SoCPi. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. Surveillance of Cerebral Palsy in Europe (SCPE). Dev Med Child Neurol. 2000;42(12):816-24.

106. Spittle AJ, Lee KJ, Spencer-Smith M, Lorefice LE, Anderson PJ, Doyle LW. Accuracy of two motor assessments during the first year of life in preterm infants for predicting motor outcome at preschool age. PLoS One. 2015;10(5):e0125854.

107. MC P. Motor Assessment of the Developing Infant. In: J D, ed. Philadelphia: W.B. Saunders,1994.

108. Burns YR, Ensbey RM, Norrie MA. The Neuro-sensory motor developmental assessment. Part 1: Development and administration of the test. Aust J Physiother. 1989;35(3):141-9.

109. Brown T, Lalor A. The Movement Assessment Battery for Children--Second Edition (MABC-2): a review and critique. Phys Occup Ther Pediatr. 2009;29(1):86-103.

110. Manuck TA, Sheng X, Yoder BA, Varner MW. Correlation between initial neonatal and early childhood outcomes following preterm birth. Am J Obstet Gynecol. 2014;210(5):426.e1-9.

111. Morgan C , Romeo DM , Chorna O , Novak I , Galea C , Del Secco S , Guzzetta A. The Pooled diagnostic accuracy of neuroimaging, general movements, and neurological examination for diagnosing cerebral palsy early in high-risk infants: a case control study. J Clin Med. 2019 Nov 5;8(11):1879.

112. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56(1):45-50.

113. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Contin Educ Anaesth Criti Care Pain. 2008;8:221-3.

114. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr. 2007;96(5):644-7.

115. Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. Crit Care. 2005;9(1):112-8.

116. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683.

117. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.

118. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710-8.

119. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal Validation of Predictive Models: Efficiency of Some Procedures for Logistic Regression Analysis. J Clin Epidemiol. 2001;54(8):774-81.

120. Ariffin SB, Midi H, eds. Robust Bootstrap Methods in Logistic Regression Model. 2012 International Conference on Statistics in Science, Business, and Engineering (ICSSBE); 2012 Sept 10-12; Langkawi, Malaysia.

121. Kruppa J, Liu Y, Biau G, Kohler M, König IR, Malley JD, et al. Probability estimation with machine learning methods for dichotomous and multicategory outcome: theory. Biom J. 2014;56(4):534-63.

122. Szymczak S, Biernacka JM, Cordell HJ, González-Recio O, König IR, Zhang H, et al. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33 Suppl 1:S51-7.

123. van Iterson M, van Haagen HH, Goeman JJ. Resolving confusion of tongues in statistics and machine learning: a primer for biologists and bioinformaticians. Proteomics. 2012;12(4-5):543-9.

124. Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. Methods Inf Med. 2012;51(1):74-81.

125. Sajda P. Machine learning for detection and diagnosis of disease. Annu Rev Biomed Eng. 2006;8:537-65.

126. Henrard S, Speybroeck N, Hermans C. Classification and regression tree analysis vs. multivariable linear and logistic regression methods as statistical tools for studying haemophilia. Haemophilia. 2015;21(6):715-22.

127. Dasgupta A, Sun YV, König IR, Bailey-Wilson JE, Malley JD. Brief review of regression-based and machine learning methods in genetic epidemiology: the Genetic Analysis Workshop 17 experience. Genet Epidemiol. 2011;35 Suppl 1:S5-11.

128. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol. 2001;54(6):603-9.

129. Brown MM. Risk Estimation Using Random forest [dissertation]. Halifax (NS): Dalhousie University, 2017.

130. Dasgupta A, Szymczak S, Moore JH, Bailey-Wilson JE, Malley JD. Risk estimation using probability machines. BioData Min. 2014;7(1):2.

131. Liu N, Koh ZX, Goh J, Lin Z, Haaland B, Ting BP, et al. Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. BMC Med Inform Decis Mak. 2014;14:75.

132. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes. 2016;9(6):629-40.

133. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of In-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med. 2016;23(3):269-78.

134. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med. 2018;36(9):1650-4.

135. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23(1):64.

136. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open. 2019;2(1):e186937.

137. Campillo-Artero C, Serra-Burriel M, Calvo-Pérez A. Predictive modeling of emergency cesarean delivery. PLoS One. 2018;13(1):e0191248.

138. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen AC, Joseph KS, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. BMC Pregnancy Childbirth. 2018;18(1):333.

139. Extremely Preterm Birth Outcomes Tool. Available from: https://www.nichd.nih.gov/research/supported/EPBO/use Accessed May 4, 2021.

140. NICHD Neonatal Research Network. Neonatal BPD Outcome Estimator. Available from: https://neonatal.rti.org/index.cfm?fuseaction=BPDCalculator.start Accessed May 4, 2021.

141. Vincer MJ, Armson BA, Allen VM, Allen AC, Stinson DA, Whyte R, et al. An Algorithm for predicting neonatal mortality in threatened very preterm birth. J Obstet Gynaecol Can. 2015;37(11):958-65.

142. Bayley N. Bayley Scales of Infant and Toddler Development Screening Test. 3rd ed. San Antonio, TX: Harcourt Assessment; 2006.

143. Vincer MJ, Cake H, Graven M, Dodds L, McHugh S, Fraboni T. A population-based study to determine the performance of the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale to Predict the Mental Developmental Index at 18 Months on the Bayley Scales of Infant Development-II in very preterm infants. Pediatrics. 2005 Dec;116(6):e864-7.

144. Weiss LG, Oakland T, Aylward GP, eds. Bayley-III: clinical use and interpretation. Cambridge: Academic Press, 2010.

145. Lee JH, Lim HK, Park E, Song J, Lee HS, Ko J, et al. Reliability and applicability of the Bayley Scale of Infant Development-II for Children With Cerebral Palsy. Ann Rehabil Med. 2013;37(2):167-74.

146. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, et al. A new and improved population-based Canadian reference for birth weight for gestational age. Pediatrics. 2001;108(2):E35.

147. The International Classification of Retinopathy of Prematurity revisited. International Committee for the Classification of Retinopathy of Prematurity. Arch Ophthalmol. 2005 Jul;123(7):991-9.

148. Bell MJ, Ternberg JL, Feigin RD, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. Ann Surg 1978; 187:1.

149. Max Kuhn KJ. Applied Predictive Modeling: New York: Springer, 2013.

150. Hosmer DW, Lemeshow S. Model-Building Strategies and Methods for Logistic Regression. Applied Logistic Regression. New York: John Wiley & Sons, Inc.; 2000. p. 91-142.

151. Liu X. Classification accuracy and cut point selection. Stat Med. 2012;31(23):2676-150.

152. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, R Core Team, 2020. Available from: https://www.R-project.org/. Accessed May 5, 2021.

153. R Studio Team. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, 2020. Available from http://www.rstudio.com/. Accessed May 4, 2021.

154. Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret Accessed May 4, 2021.

155. Liaw A WM. Classification and Regression by random Forest. R News 2002;2 (3):18-22.

156. Santhakumaran, S., et al., Survival of very preterm infants admitted to neonatal care in England 2008-2014: time trends and regional variation. Arch Dis Child Fetal Neonatal Ed, 2018. 103(3): p. F208-F215.

157. Stoll, B.J., et al., Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. JAMA, 2015. 314(10): p. 1039-51.

158. Ancel, P.Y., et al., Survival and morbidity of preterm children born at 22 through 34 weeks' gestation in France in 2011: results of the EPIPAGE-2 cohort study. JAMA Pediatr, 2015. 169(3): p. 230-8.

159. Lodha, A., et al., Do preterm infants with a birth weight ≤1250 g born to single-parent families have poorer neurodevelopmental outcomes at age 3 than those born to two-parent families? J Perinatol, 2018. 38(7): p. 900-907.

160. Statistics Canada. Census in Brief: Portrait of children's family life in Canada in 2016. Available from https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016006/98-200-x2016006-eng.cfm Accessed May 4, 2021.

161. Bronson, S.L. and T.L. Bale, The Placenta as a mediator of stress effects on neurodevelopmental reprogramming. Neuropsychopharmacology, 2016. 41(1): p. 207-18

162. Blondel B, Zuber MC. Marital status and cohabitation during pregnancy: relationship with social conditions, antenatal care and pregnancy outcome in France. Paediatr Perinat Epidemiol. 1988;2(2):125–37.

163. Colton, T., B. Janzen, and W. Laverty. Family structure, social capital, and mental health disparities among Canadian mothers. Public Health, 2015. 129(6): p. 639-47.

164. Huhtala, M., et al., Parental psychological well-being and behavioral outcome of very low birth weight infants at 3 years. Pediatrics, 2012. 129(4): p. e937-44.

165. Miceli PJ, Goeke-Morey MC, Whitman TL, Kolberg KS, Miller-Loncar C, White RD. Brief report: birth status, medical complications, and social environment: individual differences in development of preterm, very low birth weight infants. J Pediatr Psychol. 2000;25(5):353–358.

166. McIntyre, D. Taitz, J. Keogh, S. Goldsmith, N. Badawi, E. Blair. A systematic review of risk factors for cerebral palsy in children born at term in developed countries. Dev Med Child Neurol, 55 (2013):499-508

167. Schneider, R.E., et al., The Association Between Maternal Age and Cerebral Palsy Risk Factors. Pediatr Neurol, 2018. 82: p. 25-28.

168. Falster, K., Hanly, M., Banks, E., Lynch, J., Jorgensen, M., Brownell, M. and Jorm, L. Maternal age and child development outcomes at age five in Australian Aboriginal and non-Aboriginal children: a population data linkage study. Proceedings of the IJPDS Conference; Swansea, Wales. Int J Data Sci. 2017; 1(1):090. [abstract]

169. Sotiriadis, A., et al., Neurodevelopmental outcome after a single course of antenatal steroids in children born preterm: a systematic review and meta-analysis. Obstet Gynecol, 2015. 125(6):1385-96.

170. Stein, H.M., et al., Effects of corticosteroids in preterm sheep on adaptation and sympathoadrenal mechanisms at birth. Am J Physiol, 1993. 264(5 Pt 1): p. E763-9.

171. Pua, Y.H., Kang H, Thumboo J, Clark RA et al., Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation following total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc, 2020 Oct;28(10):3207-3216**.**

172. Christodoulou, E., et al., A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol, 2019 June;110:12-22.

173. Couronné, R., P. Probst, and A.L. Boulesteix, Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics, 2018. 19(1):270.

174. Synnes, A., et al., Preterm children with suspected cerebral palsy at 19 months corrected age in the Canadian neonatal follow-up network. Early Hum Dev, 2019. 136: p. 7-13.

175. The Canadian Neonatal Follow up Network. Annual Report. 2020. Available at https://cnfun.ca/annual-report/ Accessed May 4, 2021.

APPENDIX 1

Studies Providing Diagnostic Properties of Prediction of Cerebral Palsy in Preterm Infants

Literature search using the following Mesh terms; cerebral palsy AND (preterm OR premature) AND ("SENSITIVITY AND SPECIFICITY"# OR predict* OR diagnos* OR accura*). The outcome was CP either alone or as part of a composite outcome of neurodevelopmental impairment (NDI). The search was limited to studies of preterm infants with birth cohorts at or after 1990 (post surfactant era), mostly published between 2000 and 2017. After exclusion by title, 325 abstracts were reviewed, of which 41 studies were selected for full review, including some studies selected from citations references. The following 21 studies were included in this review table.

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Amplitude integrated electroencephalography (aEEG)					
Wikstrom[80] 2012 Sweden Prospective cohort	< 31 weeks 36/49 (73%) born 2005-2007 single centre	multiple records of aEEG during the first 72 postnatal hours	death or NDI at 24 months (CP, motor, cognitive, blindness, deafness) CP: Palisano[81] (GMFCS)	early aEEG recorded at 24–48 postnatal hours, is predictive of outcome with around 80% accuracy burst suppression, inter-burst intervals (IBI) & IB% predict poor outcome	for aEEG Sens, Spec, PPV, NPV & accuracy were (89, 67, 63, 91 & 76) (AUC 0.79, 95% CI, 0.65–0.93) PPV, NPV, accuracy for IBI > 6 sec were (67, 79, 74) and for IB% > 55% at 24 hrs of age were (72,80,79)
Schwindt[82] 2015 Austria Case control study	< 30 weeks & SGA 136 (47 SGA & 89 controls) single centre	multiple records of aEEG during the first two weeks of life	death or NDI at 24 months (CP, motor, cognitive) CP: Palisano	SGA infants <30 weeks had less optimal scores on early aEEG and a poorer outcome at 24 months than the AGA controls	combined aEEG score: Sens, Spec, PPV, NPV (52,80, 76,53) respectively
Cranial Ultrasound (cUS)					

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Hope[83] 2008 Canada Case control study	< 31 weeks or < 1500grams 84 (37 CP, 48 controls), born 1990-2000 single centre	cUS texture measure of white matter/ choroid plexus within first week	**CP at 24 months** CP: Palisano	quantitative early texture measures by cUS contain diagnostic information relevant to CP development	Sens, Spec (75, 69). the incidence of CP is much greater than the general population (46 vs 11%)
De Vries[84] 2004 Netherlands Prospective cohort	< 33 weeks all 429 survivors born 1990-1999 single centre	Sequential high resolution weekly cUS until term age (40 weeks)	**CP at 24 months** CP: Hagberg[85]	79% of CP cases had cUS abnormalities. Sequential cUS detected major US abnormalities in the majority of CP children CP	sequential cUS; Sens, Spec, PPV, NPV (76, 95, 48, 99). The most sensitive predictor was cystic PVL
Lacey[86] 2004 Australia Prospective cohort	< 30 weeks 203/249 (81%) born 1992-1996 single centre	cUS at day 7, 28 for IVH & LAPI before discharge (Lacey assessment of preterm infants)	**CP at 36 months** (delayedmotor development with abnormal tone)	LAPI has better diagnostic accuracy than early cUS in prediction of normal motor development or CP at 3 years of age	for cUS; Sens, Spec, PPV, NPV (44, 87, 88, 43). LAPI assessed at > 33 weeks; Sens, Spec, PPV, NPV (86, 83,57, 96)
Woodward[87] 2006 New Zealand Prospective cohort	< 31 weeks 164/167 (98%) born1998-2002 2centres	MRI at term age (81% of the cohort had MRI)	**CP or NDI at 24 months** (CP, cognitive, blindness, deafness) CP: Palisano	Moderate-severe white matter lesions on MRI were significant predictors of severe motor delay and CP after adjustment for confounders (neonatal factors & cUS findings	For CP: Sens, Spec of any white matter abnormalities (94,31) & for moderate-severe abnormalities (65,84) respectively

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Conventional (structural) MRI					
Nanba[88] 2007 Japan Prospective cohort	< 34 weeks & < 1500 grams 289/328 (88%) born 1993-2000 single centre	MRI at near term (36-43 weeks). To assess whether PVL on MRI (n=62) are predictive of CP & motor outcomes	**CP at 20 & 31 months** CP: Palisano	Lesions in the corona radiata above posterior limb of internal capsule at term MRI were predictive of motor prognosis in preterm infants with PVL	For white matter lesions; Sens, Spec, LR+, LR- (62, 87, 4.9, 0.4). For lesions in corona radiata; Sens, Spec (100, 97) respectively
Mirmiran[89] 2004 US Prospective cohort	< 30 weeks or < 1250 grams 61/99 (60%) born 1996-1999 single centre	MRI at term age compared to cUS obtained at least twice during the first 2 weeks of life	**CP at 20 & 31 months** CP: Palisano, Rosenbaum[90]	MRI predict CP better than cUS both at 20 & 31 months corrected age	At 31 months corrected age; MRI Sens, Spec, LR+, LR- (86, 89, 10, 0.1) compared to cUS (43, 82, 2, 0.7) respectively
General movements assessment (GMA)					
Ferrari[91] 2002 Italy Prospective cohort	< 37 weeks with significantly abnormal cUS 84/93 (90%) single centre	cramped sync hronized GM from birth until 56-60 wks vs neurological exam	**CP at 24-36 months** CP:Ellenberg[92]	Consistent & predominant cramped s ynchronized GM specifically predict CP. The earlier this appears, the worse is the later disability	Cramped synchronized GM predicts CP better than ultrasound (AUC 0.97 vs 0.88) and neurologic exam; Sens, Spec (79, 100) vs (8)
Study	**Population**	**Exposure/ Intervention**	**Outcome**	**Findings**	**Diagnostic properties of model**
Romeo[93] 2008	< 37 weeks	GMA "Fidgety	**CP or NDI at 24 months**	Combining the 2 methods is more	For single assessment, GMA is better

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Italy Prospective cohort	903/925 (98%) born 2000-2004 single centre	movements" combined to neurologic exam(HINE)[94] at 3 months corrected age	(CAT-CLAMS[95]quotient below 70 but no CP) CP: Hagberg	effective than single assessment in predicting outcome particularly for discriminating unilateral and bilateral CP	predictor of CP compared to HINE; Sens, Spec (98,94) for GMA vs (96,87) for HINE score < 57
Oberg[96] 2015 Norway Prospective cohort	<33weeks or < 1500 grams 87/173 (50%) born 2002-2010 Single centre	"Fidgety movements" at 3 months in a routine clinical setting	**CP at 24 months** CP: Palisano	Absence of "Fidgety movements" at 3 months corrected age predict CP & motor outcome at 2 years of age	Sens, Spec, LH+, LH (90,90,8.7,0.1) The NPV 99% & PPV 53%
Adde 2010[98] Norway Prospective cohort	13/30 (43%) high risk infants (23-42 weeks) born 2002-2004	video analysis of GMA at 10-15 weeks (1 record to assess fidgety movements)	**CP at 5 years** CP: Palisano	Variability of centroid motion at 10-15 weeks , with assessment of fidgety movements, predict CP at 5 years	Sens, Spec 85 and 71 Specificity increased to 88% when combined with variables of the amount of motion
De Bock[97] 2017 Germany Prospective cohort	< 33 weeks 122/256 (48%) born 2007-2009. Single centre	GMA at 1 & 3 months in a routine clinical setting	**CP or MDI/ PDI ≤70 at 24 months** CP: Palisano	Definitely abnormal GM at 3 months corrected age identified all children with CP at 2 years	Definitely abnormal GM were predictors of atypical outcome (Sens, Spec; 56, 82)

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Stahl, 2012[99] Norway Prospective observational study	15/82 (CP 18%) infants (preterm or term) with 1-2 videos at 10-18 weeks corrected age	applying computer vision-based (optical flow) movement assessment and statistical pattern recognition	**CP at 2-5 years** CP: not specified	Early detection of CP can be done using SVM machine learning of 3 movement patterns for classification of CP (The simple features (relative frequency and absolute motion distance) comprised a higher discrimination than the feature based on wavelet decomposition of the signal.	Accuracy of 93.7% +/- 2.1, Sens, Spec of 85 and specificity of 95
Combined conventional MRI & GMA					
Constantino[16] 2007 US Prospective cohort	<32wks &<1500 g 102/130 (78%), born 1996-1999 single centre	combined MRI at term, GMA & behavioral assessment (NAPI)[100]	**CP at 18 months** CP: Palisano, Amiel-Tison[101]	All tests NPV 90-97%. For Spec & accuracy; MRI was superior (91& 84), GM at 52 wks was better than at 36wks. Sensitivity increased with NAPI + MRI	for combined MRI & NAPI; Sens, Spec, PPV, NPV & accuracy (80, 81, 36, 97, 80)
Skoild[102] 2013 Sweden Prospective cohort	< 27 weeks all 53 infants born 2004-2007 population based study	Combined MRI at term age & GMA at 3 months	**CP and/or abnormal motor development at 30 months** Tests:SCPE[103], Palisano	Moderate –severe white matter injury on MRI predicts CP better than abnormal GMs. When combined, increase GM specificity to 100% but did not affect Sensitivity	Abnormal MRI vs abnormal GM; Sens, Spec, PPV, NPV; (100, 98, 80, 100) vs (50, 92, 33, 96). Combining both; Sens, spec, PPV, NPV (50,100, 100,96)

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Others					
Broitman[17] 2007 US Retrospectiv e cohort	ELBW(<1000 g) 2103/2750 (76%) born 1998-2001 19centres NICHD	Clinical model compared to cUS (early at 28 days & late at 36 weeks)	NDI at 18-22 months (CP, cognitive, blindness, deafness) & independently walk/feed Tests: Amiel-Tison	The clinical models were better predictors than early and late cUs for NDI (AUC 0.68 vs 0.58 and 0.57, p<0.001). Isolated cUS findings were poor predictors of CP. Only PVL at 36 weeks (OR 5.2 (2.8–9.6)) and VP shunt (OR 3.7 (1.8-7.8)) were predictive of CP	Improvement in the predictive ability (AUC) for mental developmental index<70 (0.72 vs 0.69), CP (0.78 vs 0.72) and independent walking (0.79 vs 0.74) for the cUS-36/"All" clinical model as compared to the cUS-28/"Early" clinical model.
Tyson[18] 2008 US Retrospectiv e cohort	22-25 weeks 4165/4446 (94%) born1998-2003 19 centres NICHD	only gestational age for providing intensive care at the edge of viability. To develop a predictive model of death, death with any NDI or death with severe NDI for counseling	death or NDI at 18-22 months (CP, cognitive, blindness, deafness) CP: Palisano	Each 100 grams increased birth weight, female sex, antenatal steroids, singleton were each associated with reductions in risks of death and death/NDI similar to the reductions with a 1-week increase gestation	The five-factor model provided for death; AUC 0.75 (0.74–0.77) and for death or NDI; AUC 0.75 (0.73-0.77)

Study	Population	Exposure/ Intervention	Outcome	Findings	Diagnostic properties of model
Spittle[104] 2015 Australia Prospective cohort from a RCT of a preventive care program to improve development	< 30 weeks 97/138 born single centre	Combined 2 motor tests (AIMS[105]/ NSMDA)[106] over the first year of life	**CP and motor impairment at 48 months** CP: Palisano & Movement Assessment Battery for Children (2)[107]	Although false positives were common, CP was most accurately predicted by NSMDA at 12 months whereas AIMS at 4 months provided the best accuracy for motor impairment	Combining both the NSMDA and AIMS provided the best accuracy at 4 months, although results were similar at 8 and 12 months Combined tests Sens, spec, PPV & NPV for CP (83, 93, 45, 99) and predictive accuracy of 92 (84, 97)
Manuck[108] 2014 US Secondary analysis of RCT of antenatal magnesium sulfate	< 32 weeks 1771/1954	neonatal diagnoses prior to initial hospital discharge	NDI (CP or MDI/PDI< 2SD) at 24 months CP: Palisano	models of individual neonatal morbidities were moderately predictive of NDI after controlling for GA, maternal education maternal race, use of tobacco, alcohol or drug, fetal sex, magnesium & chorioamnionitis	best model for NDI had AUC of 0.68 (0.65-0.71). Combinations of 2, 3 & 4 morbidities did not improve NDI prediction
Morgan et al, 2019[109] Italy (3 sites) retrospective case control study born between 2002 - 2016	441 preterm and term high risk infants (147 CP, 147 mild disability, 147 controls)	HINE scores at 3 months+ early brain imaging (MRI or US) + Absent fidgets on GMA	**CP at 24 months** CP: Palisano	the pooled analysis with the 3 predictors provided the highest AUC compared to any individual predictor (AUC 0.99 vs 0.85, 0.96, 0.96 respectively)	AUC 0.99 98.74% of children were correctly classified, Sensitivity for detecting CP was 98%, and specificity was 99% (PPV 98.56; NPV 98.84)

APPENDIX 2

Table S1: Characteristics of Population Dataset

Variable	Type	Code	Definition
Sociodemographic Maternal Variables			
Maternal age in years	Continuous	matAge	maternal age in years
Hollingshead socioeconomic status at birth	Categorical	ses	Class I to V
Married or common law	Binary	marriedCL	married or common law vs single parent
Urban accommodation	Binary	urban	urban vs rural accommodation
Prenatal Variables			
Primigravida	Binary	gravid1	primigravida
Multiparity	Binary	multipara	More than one previous delivery
Abortion/miscarriages	Binary	aborMisc	previous abortion/miscarriages
Previous stillbirths	Binary	prvStb	previous stillbirth
Maternal smoking during pregnancy	Binary	smk	any smoking during pregnancy
Maternal substance use during pregnancy	Binary	substUse	any substance use (illicit or non illicit)
Maternal antidepressants*	Binary	antidepr	treatment for anxiety/depression during pregnancy
Maternal psychiatric disorder*	Binary	psych	psychiatric disorder during pregnancy
Maternal treatment for diabetes*	Binary	diabAny	any treatment for diabetes during pregnancy
Maternal hypertension*	Binary	hyptAny	gestational or pre-existing hypertension
Chorioamnionitis /funisitis *	Binary	chorioFuni	histologic chorioamnionitis or funisitis
Prelabor premature rupture of membranes	Binary	pprom	prolonged rupture of membranes >18 hours
Maternal antibiotics*	Binary	ipAbx	maternal intrapartum antibiotics

Variable	Type	Code	Definition
Mother colonized with group B streptococci	Binary	gbsPos	Maternal group B streptococci colonization during pregnancy
Antepartum hemorrhage*	Binary	apHemo	Any antepartum hemorrhage
Placental abruption	Binary	abruptio	placental abruption
Any tocolytic use*	Binary	tocolyt	any tocolytic
Maternal indomethacin	Binary	matindo	indomethacin for tocolysis
Fetal growth restriction	Binary	fetIUGR	fetal growth restriction by ultrasound
Fetal distress	Binary	fetDistr	fetal abnormal heart tracing or fetal acidosis
Perinatal Variables (including intrapartum and the first 6 postnatal hours)			
Antenatal steroids	Categorical	anSteroid	(0 , none),(1 ,<24 hrs}, (2 ,24-47 hrs},(3 ,48-167 hrs},(4 ,>= 168 hrs}
Optimal antenatal steroids	Binary		> 24 hours prior to delivery
Intrapartum magnesium sulfate	Binary	mgsulf	intrapartum magnesium sulfate
Gestational age in weeks#	Continuous	ga	gestational age in weeks
Extreme low gestational age	Binary	elgan	gestational age < 26 weeks
Birth weight in grams	Continuous	bw	birth weight in grams
Extreme low birth weight	Binary	elbw	birth weight <1000 grams
z-scores of weight for age	Continuous	bwz	z scores of birth weight based on Canadian growth curves (Kramer)
Small for gestational age	Binary	sga	< 10^{th} centile based on Canadian growth curves (Kramer)
Male sex	Binary	sexY	male vs female
Outborn delivery	Binary	outborn	outborn vs inborn
Moderate to severe birth depression	Binary	modsevAsphyx	receipt of positive pressure ventilation or resuscitation
1-minute Apgar score	Continuous	apgar1	Apgar score at 1 minute
5-minute Apgar score	Continuous	apgar5	Apgare score at 5 minutes

Variable	Type	Code	Definition
Chest compression/epinephrine at delivery	Binary	resusDel	resuscitation at delivery
Delivery by Caesarean section	Binary	cs	Caesarean vs vaginal delivery
Admission temperature (degrees Celsius)	Continuous	admTemp	NICU admission temperature (degrees Celsius)
Admission hemoglobin (g/L)	Continuous	admHgb	NICU admission hemoglobin (g/L)
Hypotension on admission	Binary	admBPLow	mean blood pressure less than gestational age at birth
Postnatal Variables			
Lowest hemoglobin in the first 24 hours	Continuous	lowestbp	lowest hemoglobin during the first 24 hours
Neonatal Insulin therapy	Binary	neoInsulin	severe hyperglycemia requiring Insulin
Neonatal hypoglycemia	Binary	neoHypoglyc	severe hypoglycemia <1.67 mmol/L
Neonatal anemia	Binary	neoAnemia	neonatal anemia
Neonatal thrombocytopenia	Binary	neoThrombpen	thrombocytopenia (<100,000)
Cystic white matter lesions^	Binary	cystWMD	cystic PVL or porencephaly
Parenchymal echodense lesions	Binary	echodensWMD	parencyhmal hemorrhage or ischemia
Severe intraventricular hemorrhage^	Binary	ivhGrade34	grade 3, 4 IVH
Posthemorrhagic hydrocephalus^	Binary	phhydro	hydrocephalus following severe IVH
Ventriculoperitoneal shunt for hydrocephalus*	Binary	phhshunt	hydrocephalus requiring shunt
Necrotizing enterocolitis	Binary	nec	NEC ≥ stage2 Bell's
Patent ductus arteriosus (PDA)	Binary	signifPDAp	hemodynamically significant PDA

Variable	Type	Code	Definition
Persistent pulmonary hypertension	Binary	pfcnew	pulmonary hypertension of newborn
Severe Retinopathy of prematurity (ROP)	Binary	sevROP	severe ROP ≥ stage 3
Intervention for ROP	Binary	ropTx	surgery or Bevacizumab intravitreal injection
Surfactant therapy*	Binary	surfact	surfactant for respiratory distress syndrome (RDS)
Severe hyaline membrane disease	Binary	sevHMD	severe RDS requiring invasive mechanical ventilation
Nasal ventilation	Binary	ncpap	nasal continuous positive airway pressure (CPAP)
High frequency oscillatory ventilation	Binary	hfov	High frequency oscillatory ventilation
Hours on tracheal ventilation	Continuous	durETT	duration of tracheal (invasive) mechanical ventilation in hours
Pneumothorax	Binary	pneutx	pneumothorax requiring drainage
Pulmonary Hemorrhage	Binary	pulmHemo	pulmonary hemorrhage
Cystic bronchopulmonary dysplasia	Binary	cystBPD	cystic BPD
Dexamethasone	Binary	neoDexa	systemic steroids for BPD
Oxygen use at discharge from the nursery*	Binary	homeO2	discharge on home oxygen
Neonatal septicemia	Binary	neoSeptic	positive blood culture
Clinical (culture negative) sepsis	Binary	neoClinSepsis	infection treated with antibiotics > 5 days
Systemic infection	Binary	noeSystinf	pneumonia, cellulitis, bacteremia, urinary infection
Any cardiopulmonary resuscitation*	Binary	cpr	any CPR during hospital stay
Inotropes*	Binary	inotrop	neonatal inotropes
Inhaled Nitric oxide*	Binary	iNO2	inhaled nitric oxide therapy

Variable	Type	Code	Definition
Hours on total parenteral nutrition	Continuous	durTPN	duration of parenteral nutrition in hours
Major surgery	Binary	nepSurgery	major neonatal surgery
Prophylactic indomethacin	Binary	prphylindo	neonatal prophylactic indomethacin
Total length of stay (days)	Continuous	neoLOS	total hospital stay in days in all nurseries
Outcome Variables			
Corrected age at latest assessment	Continuous	ageseen	post term age seen at last assessment in weeks
Normal at latest assessment	Binary	normal	no neurodevelopmental impairment (no CP
Cerebral palsy (CP)	Binary	cp	(0 no CP},(1, CP}
Cerebral palsy severity	Categorical	cpstage	CP GMFCS stages 1 to 5

*Atlee database code

Abbreviation: BPD (bronchopulmonary dysplasia), CP (cerebral palsy), CPAP (continuous positive airway pressure), CPR (cardiopulmonary resuscitation), GMFCS (gross motor functional classification system), NEC (necrotizing enterocolitis), NICU (neonatal intensive care unit), PDA (patent ductus arteriosus), PVL (periventricular leukomalacia), RDS (respiratory distress syndrome)

Confirmation of gestational age is determined according to the following hierarchical order:

1. Conception dating, if mother was receiving fertility treatments;
2. The last menstrual period, if it corresponds to ultrasound dating within 10 days;
3. Ultrasound dating, if it was >10 days difference from the last menstrual period or no dates were known;
4. Physical examination of the infant at birth, if none of the three preceding estimates were available

^ Routine cranial ultrasound screening of all preterm infants is standard of care at the IWK. This includes serial cranial ultrasound imaging with the initial screening is done between 3-7 days after birth, then at 2 and 6 weeks after birth and finally at term equivalent age. If an abnormality is identified, more imaging is performed as clinically indicated. Brain MRI is done in a selected subgroup of infants with severe abnormality identified on routine cranial ultrasound as per the discretion of the treating physician. Reporting of abnormal findings on neuroimaging includes the type, site

(unilateral or bilateral) and the extent of brain injury. However, coding of abnormal neuroimaging in the PFUP database includes the worst finding (type) and laterality.

Table S2: Missing Values in Population Dataset

	Missing Values	Number of Observations	% Missing
Continuous Variables			
Apgar score at 1 minute	8	769	1.0
Apgar score at 5 minutes	8	769	1.0
Z scores of birth weight	2	775	0.2
Admission hemoglobin	8	769	1.0
Admission temperature	18	759	2.3
Lowest hemoglobin during first 24 hours	5	772	0.6
Days of parenteral nutrition	14	763	1.8
Hospitalization days	7	770	0.9
Binary/Categorical Variables			
Married or common Law	10	767	1.3
Optimal antenatal steroids (>24 hours prior to delivery)	1	776	0.1
Chest compression /epinephrine	4	773	0.5
Hypotension on admission	15	762	1.9
Prophylactic indomethacin	1	776	0.1
cystic bronchopulmonary dysplasia	1	776	0.1
Socioeconomic Status	88	689	**11.3**
Smoking	40	737	**5.1**

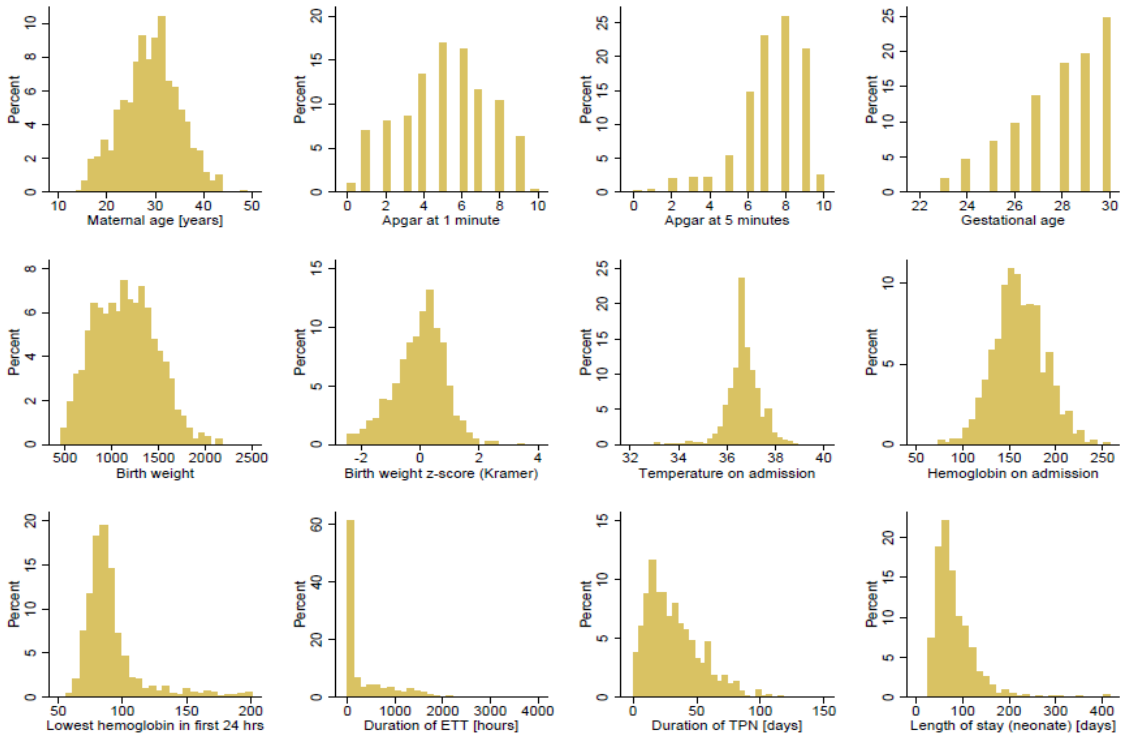

Figure S1. Distribution of Continuous Variables in the Population Dataset

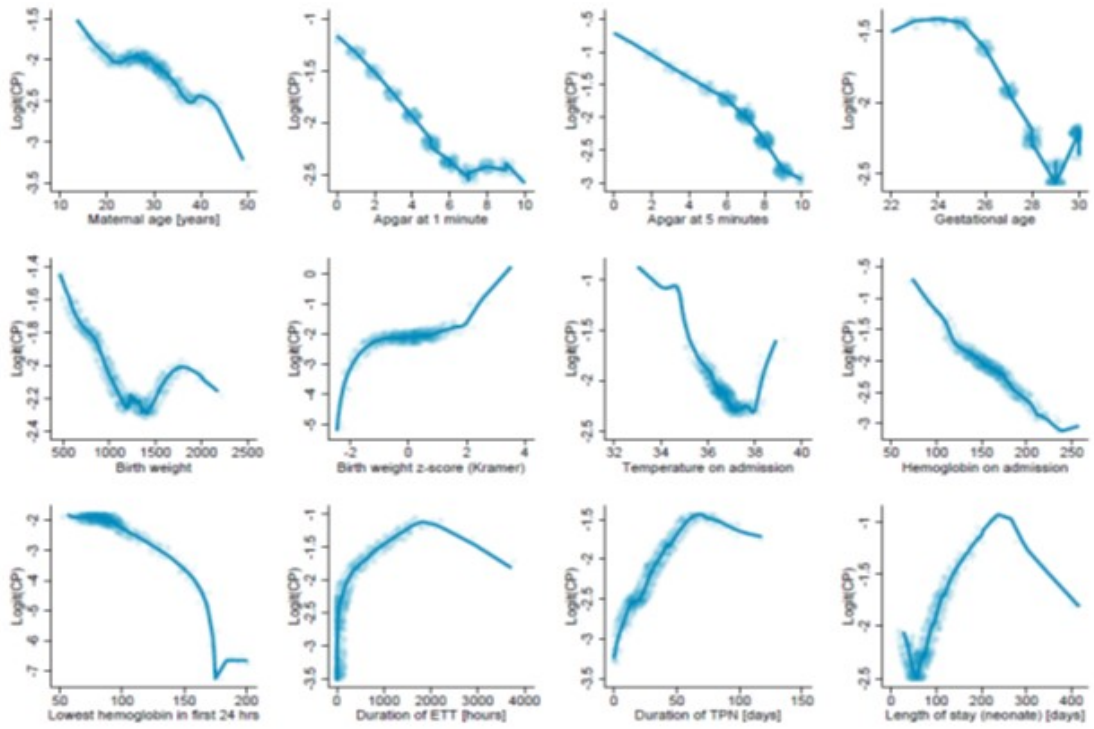

Figure S2. Assessment of the Linear Relation Between Continuous Variables and Logit of CP in the Population Dataset

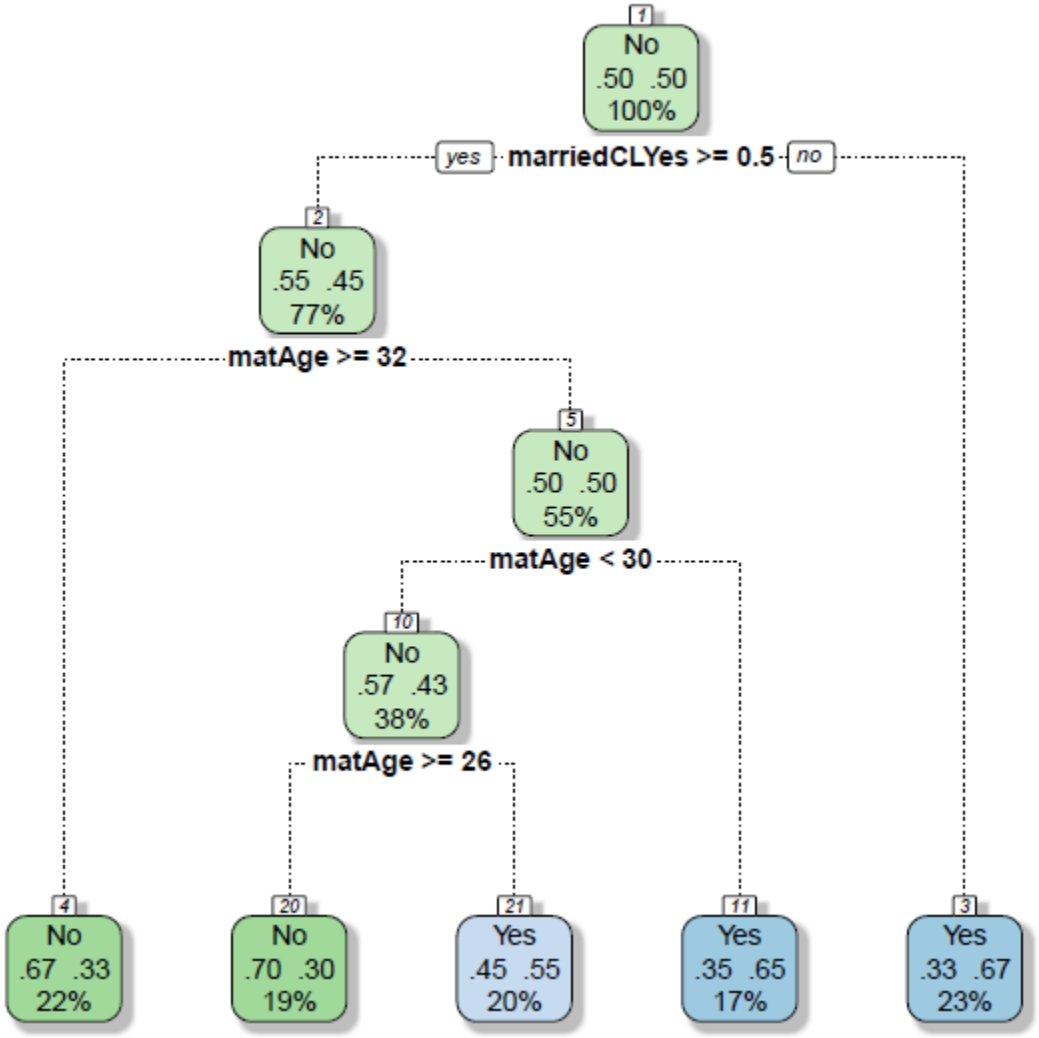

Figure S3.a. Classification Tree of the Prenatal Model

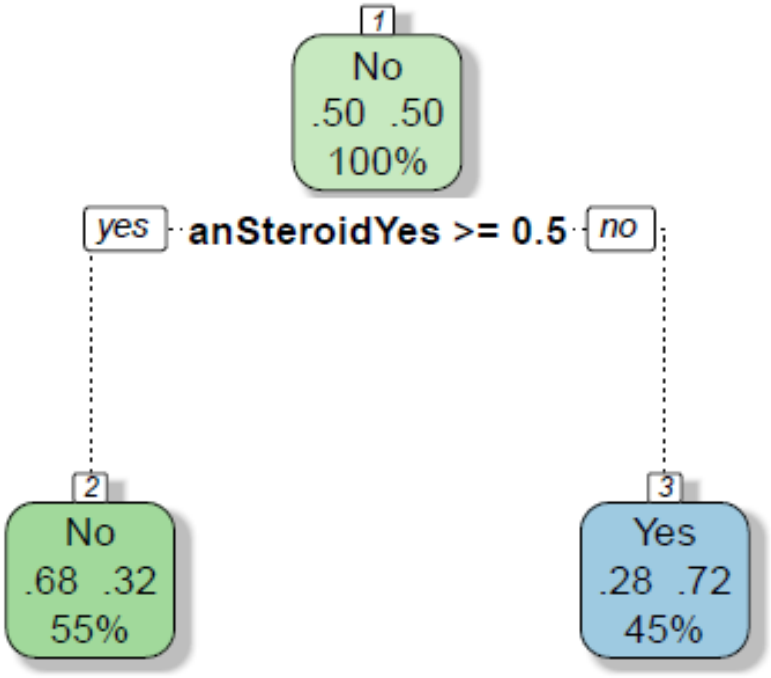

Figure S3.b. Classification Tree of the Prenatal-Perinatal Model

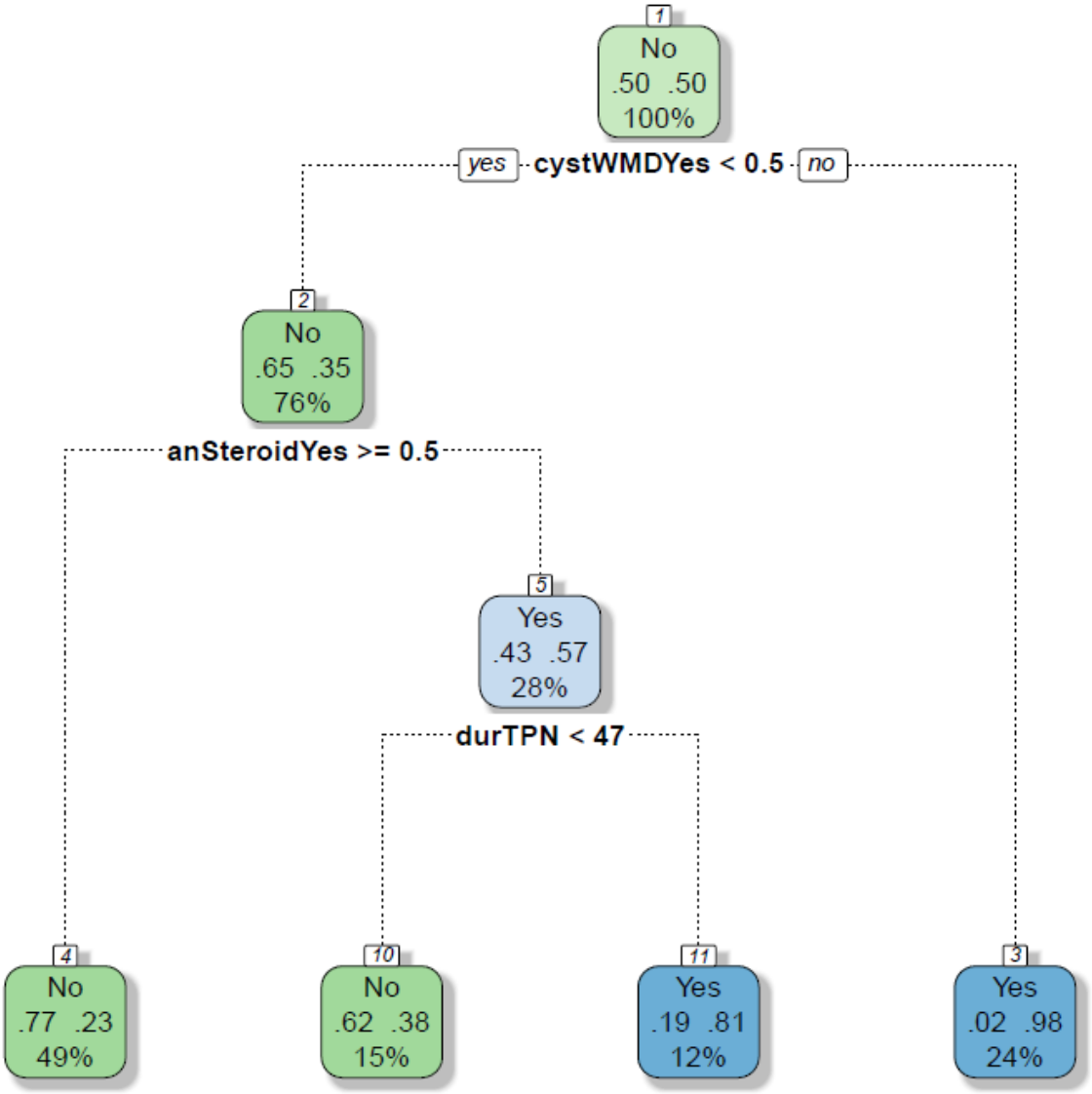

Figure S3.c. Classification Tree of the Full Mode