The Effect of Neurodevelopmental Impairment Definition on Incidence Rates among Very Preterm Infants

by

Matthew David Haslam

B.A., Washington University in St. Louis, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
in
The Faculty of Graduate and Postdoctoral Studies
(Population and Public Health)

THE UNIVERISITY OF BRITISH COLUMBIA
(Vancouver)

October 2016

Abstract

Background: Various criteria are used to define severe neurodevelopmental impairment (SNI) and the effect of definition is rarely reported.

Objective: To examine the impact of changes in SNI definition on incidence rates of SNI and the association between risk factors and SNI.

Methods: We included infants (n=2187) born <29 weeks gestation between April 2009 and September 2011, who were admitted to a Canadian Neonatal Network Neonatal Intensive Care Unit (NICU) and assessed at 18-21 months corrected age by the Canadian Neonatal Follow-Up Network (CNFUN). Incidence rates of SNI were calculated for 7 commonly used definitions identified in the literature. Logistic regression was performed to identify risk factors for SNI using the definitions which yielded the highest and the lowest incidence rate of SNI. Adjusted odds ratios (AOR) and 95% confidence intervals (CI) were estimated for risk factors significantly associated with SNI.

Results: SNI definitions were composed of six common criteria: cerebral palsy severity using the Gross Motor Function Classification System (GMFCS), motor, language, and cognitive Bayley-III scores, and visual or hearing impairment. SNI incidence ranged from 3.5% to 14.9% (highest vs lowest rate ratio 4.29; 95% CI: 3.37, 5.47). The definition yielding the highest incidence included at least one of: GMFCS score 3-5, Bayley-III motor, language, or cognitive score <70, bilateral visual impairment, or use of hearing aids or cochlear implants. The definition yielding the lowest incidence included at least one of: GMFCS score 4-5, Bayley-III language or cognitive score <55, or bilateral visual impairment. The associations between risk factors and SNI varied depending on the SNI definition used. Maternal ethnicity, employment status, antenatal steroid treatment, and gestational age at birth were inconsistent in the significance of their associations with SNI. Maternal drug use, infant male sex, score of neonatal acute physiology >20, late onset sepsis, bronchopulmonary dysplasia, and intraventricular hemorrhage were consistently associated with SNI, irrespective of the SNI definition used, although the strength of these associations varied.

Conclusions: Criteria used to define SNI significantly influence SNI incidence and the associations between risk factors and SNI. A standardized definition of SNI would facilitate scientific communication and spatio-temporal comparisons.

Preface

Under the guidance of Dr. Anne Synnes, Dr. Sarka Lisonkova, and Dr. K.S. Joseph, I was directly involved in the conception and direction of this project for the entirety of its duration. Results included in this thesis have been presented at national and international conferences (Canadian National Perinatal Research Meeting February 10th, 2016; Pediatric Academic Society annual meeting May 1st, 2016; Society for Pediatric and Perinatal Epidemiologic Research annual meeting June 20th, 2016). Due to confidentiality of individual level data, all analyses presented in this document were performed at my request by statisticians at the coordinating MiCare site, University of Toronto/Mount Sinai Hospital, in Toronto, ON, which houses data for the Canadian Neonatal Network (CNN) and Canadian Neonatal Follow-Up Network (CNFUN).

Contributors:

- Dr. Anne Synnes and Dr. Sarka Lisonkova – provided direct supervision of the project and were integral in the development of the protocol and overseeing analysis and writing.
- Dr. Paige Church and Dr. Dianne Creighton – provided recommendations as experts in their fields of clinical practice and research, as well edits on abstracts, posters, manuscripts, and this thesis.
- Dr. K.S. Joseph – served on the thesis review committee and provided supervision for the project.
- The Maternal Infant Care Research Coordinating Site, which houses the Canadian Neonatal Network™ (CNN) and the Canadian Neonatal Follow-Up Network (CNFUN), provided access to the data used in this study and Mr. Junmin Yang performed all statistical analyses presented in this thesis based on my requests for specific analyses.

Ethics Approval: This study was approved by the British Columbia Children & Women's Hospital Research Ethics Board: H15-00605.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOR	adjusted odds ratio
Bayley-III	Bayley Scales of Infant and Toddler Development, 3rd Edition
BPD	bronchopulmonary dysplasia
BW	birth weight
CI	confidence interval
CNN™	Canadian Neonatal Network
CNFUN	Canadian Neonatal Follow-Up Network
CP	cerebral palsy
CRIB	Critical Risk Index for Babies
GA	gestational age
GMFCS	Gross Motor Function Classification System
IVH	intraventricular hemorrhage
MiCare	CIHR Team in Maternal and Infant Care
NEC	necrotizing enterocolitis
MNI	moderate neurodevelopmental impairment
NICU	neonatal intensive care unit
PDA	patent ductus arteriosus
PVL	periventricular leukomalacia
RR	rate ratio
ROP	retinopathy of prematurity
SGA	small for gestational age
SNAP-II	Score for Neonatal Acute Physiology, 2nd Edition
SNI	severe neurodevelopmental impairment
VPT	very preterm infant

Acknowledgements

I would like to thank my supervisors Dr. Anne Synnes, Dr. Sarka Lisonkova, and Dr. K.S. Joseph for their constant aid throughout this project, without whom this project would not have been possible. Their patient guidance has shaped me as a researcher and I am confident that they will prove to be lifelong inspirations as I move forward in my career.

Thanks to CNN and CNFUN for providing such a wealth of data, and to Dr. Prakesh Shah and Mr. Junmin Yang in particular for their support for the analyses in this project.

Thanks to my friends and family for pushing me to be academically driven and capable throughout my educational career, and also teaching me the value of practicing compassionate humanism in all of life's arenas.

Finally, a special thanks to all of the parents and their infants who participated in CNN, CNFUN, and MiCare data collection, whose data was used in this study, and who courageously live through the challenges associated with very preterm birth every day.

1. Introduction

1.1 Incidence and impact of prematurity

The rate of preterm birth (gestational age <37 completed weeks) has been steadily increasing since the early 1980's.(1) In Canada, the rate of preterm birth increased by 21% from 6.6 to 8.0 per 100 live births between 1990 and 2008, and then declined to 7.6 per 100 live births in 2011.(2,3) The consequences of preterm birth are serious and far-reaching. Prolonged hospital admissions and high technology life-saving support in the neonatal intensive care unit lead to elevated neonatal health care costs.(4) The annual cost to the healthcare system associated with preterm birth is estimated to be $587.1 million in Canada, and is even higher in the United States – between $5.8-$26.2 billion per year.(5–9) A large portion of the economic burden can be attributed to the 5–15% of preterm births that are considered to be "very preterm" (VPT).(1) In this report, infants delivered before 29 weeks' gestational age are considered to be VPT, but the gestational age cutoff for classifying infants as VPT is variant in the literature. According to one study, of all the preterm births in Canada only 6.7% were born before 28 weeks, however this group was responsible for 21% of the annual national cost associated with preterm birth ($123.3 million).(5) It is estimated that the average healthcare costs for a VPT infant, from pregnancy through the neonatal period, ranges from $54,098 to $90,123 in Canada.(5,10) Very preterm births therefore have a significant impact on the healthcare systems in Canada and other industrialized countries.

1.2 Mortality

Preterm birth is the second leading cause of infant mortality in Canada, after “congenital malformations & chromosomal abnormalities.”(11) Canada has fallen to 28th of 36 OECD countries in terms of infant mortality, in large part due to the high incidence of preterm associated mortality.(12) The main driver of high mortality rate among preterm infants is the most immature and vulnerable group – very preterm infants. Survival rates improve dramatically with increasing length of gestation (13–15), but mortality rates at the lower limit of viability (22 weeks GA) can reach 94%.(16) Furthermore, for many of the VPT infants who do survive, significant barriers to normal health and development remain.

1.3 Short Term Morbidity

Very preterm infants require admission to the neonatal intensive care unit (NICU) and usually require resuscitation at birth. Major organ systems such as the pulmonary, cardiovascular, gastrointestinal and central nervous system are immature and vulnerable in the neonatal period.(15) Major neonatal morbidities affecting short-term and long-term health outcomes include bronchopulmonary dysplasia (BPD) (17), patent ductus arteriosus (PDA) (18), necrotizing enterocolitis (NEC) (19), brain injury such as intraventricular hemorrhage (IVH) and periventricular leukomalacia (20,21) and sepsis.(22) Each of these conditions can be life threatening, and directly affect an infant’s ability to thrive. Birth at early gestation is associated with very low birth weight; however, preterm infants that are also small-for-gestational age (SGA) have additionally increased

risks for adverse health outcomes.(23) Severity of illness soon after birth can be measured using indices such as the Critical Index for Babies – 2nd Edition (CRIB-II), and the Score for Neonatal Acute Physiology – 2nd Edition (SNAP-II). These indices are intended to quantify the illness severity and the risk of death at the time of admission to the NICU.(24–26) Infants with lower CRIB-II scores or higher SNAP-II scores are more likely to have severe adverse short-term outcomes including death.(25–28) The initial severity of illness and complications acquired over the course of the infant's stay in the NICU can also increase the risk for long-term disability. The various short-term morbidities listed above have all been associated with adverse neurodevelopmental outcome at 18-24 month follow-up.(15,17–22,24,29)

1.4 Long-Term Morbidity

Infants born very preterm have worse long-term prognosis than infants born at term (14); between 3.5% and 25% of VPT infants have profound developmental impairments detected by 18-24 months corrected age.(30,31) Major neurodevelopmental problems include cerebral palsy and diminished motor function (20), and impaired cognitive and language development.(32–34) In addition, sensory impairments such as hearing and vision loss may also affect functional ability for preterm survivors.(35,36) Early identification of infants with a high likelihood of long-term neurodevelopmental impairment is important for early intervention and initiation of rehabilitation therapy.(37)

Furthermore, it is important to remember that even though infants born very preterm are at significantly higher risk of developing neurodevelopmental impairments than term-born infants, impairment or disability is not an inevitable consequence of preterm birth. In fact, even at the lower limits of viability some children do very well with no evident sequelae of prematurity.(38–40) The fact that some very preterm infants develop normally provides hope for improving interventions to bolster the likelihood of functionally normal VPT survivors. Studying children who develop normally and those who are neurodevelopmentally impaired can provide insight into which risk factors diminish or facilitate normal growth and development. To screen for possible impairments, clinicians and researchers perform prospective longitudinal assessments of VPT survivors to assess their developmental progress and identify warning signs of developmental disorders. Signs and symptoms of severe neurodevelopmental impairment can be observed at the corrected postmenstrual age of 18-24 months. This time period is the most common period used to classify an infant as having moderate or severe neurodevelopmental impairment.(41) Classifications of normal development, moderate neurodevelopmental impairment (MNI) and severe neurodevelopmental impairment (SNI) are defined according to specific criteria, and are based upon results of specific examinations.

A multiplicity of causes for impairment exists, and identifying risk factors associated with neurodevelopmental impairment is important for guiding clinical practice and epidemiological research aimed at reducing the incidence of severe

impairment and death in the very preterm population. Risk factors for adverse neurodevelopmental outcomes in very preterm infants can be identified before or during pregnancy, at birth, and in the neonatal and post-neonatal periods (Appendix: Figure 2).

1.4.1 Cognitive and language impairment

VPT infants have a higher risk of developmental impairment in cognition and language compared with their full term counterparts. Rates of severe cognitive and language impairment were 2.3% and 2.2%, respectively, in a cohort of VPT survivors from the Australian and New Zealand Neonatal Network which defined severe impairment as a score three standard deviations below the mean on the Bayley-III scale (<55).(31) In the United States, the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network (NICHD), which defined severe cognitive impairment as a Bayley-III cognitive score more than two standard deviations below the mean (<70), yielded a severe cognitive impairment rate of 9.3% among VPT surviving infants.(42) Two parallel studies from the UK and Australia (the BOOST-II trial) found 28.4% of infants had either a cognitive or language score one standard deviation below the mean (<85), which the authors considered to be a severe impairment.(43) Although variations in definitional criteria makes estimating true incidence challenging, these studies nonetheless highlight the problem of relatively frequent cognitive and language impairment in the very preterm population.

The Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III) is a developmental test administered between 1-42 months of age and is comprised of three components with composite scores: Cognitive, Language, and Motor.(33,44) The test was developed using a normative population of toddlers with 90% displaying typical development and 10% with established developmental impairment.(44) The cognitive scale is comprised of 91 items that assess an infant's sensorimotor development, exploration and manipulation, object relatedness, concept formation, and memory. The language scale has 97 items that assess receptive communication (comprehension) and expressive communication (vocabulary). Not all of the items are administered to the children, however. The test has item sets A to Q and the starting point depends upon the child's adjusted age and ability to complete the first 3 tasks in a row within an item set. The assessment then continues until the tasks become too complex for the child to complete, demonstrated by failing 5 consecutive tasks. For example, toddlers at 18 months corrected age typically start at section K or L whereas young children at 24 months would start at section M.(45) Each composite component is rated on a numeric scale with a standardized mean score of 100 and standard deviation of 15.(44) Lower scores indicate poorer performance. Though low scores usually reflect developmental delay or impairment, occasionally scores will not represent typical performance (e.g. a "bad day", feeling unwell) or will be affected by behavioural issues. When interpreted by a clinically trained tester, the Bayley-III is a useful tool that allows

parents and physicians to discuss early intervention strategies geared towards improvement.

1.4.2 Motor impairment

As with the cognitive and language portions of the Bayley-III, infants who score poorly on the motor function portion may be flagged for further assessment because of concerns related to cerebral palsy or other motor impairment. The Bayley-III motor composite score is comprised of 138 items that test fine and gross motor function using object manipulation and observing dynamic movement while walking or crawling.(44) Again, the child will start at the age-appropriate starting point and end when they fail several tasks in a row. Rates of severe motor function impairment have been reported to be approximately 8.3% in the very preterm population when using a cut-off of greater than 2 standard deviations (SD) below the mean (<70) (42), and 2.9% when defining severe impairment as greater than 3 SDs below the mean (<55).(46) Children with severe motor impairment may have difficulty with general movement such as ambulation, and they may also experience a significant reduction in dexterity and fine motor skills. Some, but not all, who have severe motor impairments will be diagnosed with cerebral palsy, which has a range of severity and impact on functional ability.

Cerebral palsy (CP) is “a group of disorders of the development of movement and posture, causing activity limitation, that are attributed to non-progressive disturbances that occurred in the developing foetal or infant

brain."(47) Cerebral palsy encompasses a wide range of severity, as this term includes a number of different types of impairment. Only one limb may be affected (monoplegia), or just one side of the body (hemiplegia); however, CP is sometimes more extensive, affecting both the arms and legs (diplegia or quadriplegia).(48) Furthermore, the location of brain injury differentially affects the type of movement disorder. The most common type of CP in the VPT is spastic CP, a condition characterized by the co-stimulation of motor neurons that leads to tightened, spastic muscles. This form of CP represents roughly 75% of people with CP, and results from damage to the motor cortex.(48) Spastic diplegia is also the form most commonly associated with preterm birth.(49) In contrast, more extensive damage to the basal ganglia can result in a less common form of CP called athetoid/dyskinetic CP, and damage to the cerebellum can result in the least common form, ataxic CP (occurring in 25% and 5-10% of individuals with CP, respectively). Both athetoid and ataxic CP can cause difficulty in walking or sitting, and sometimes speaking.

Furthermore, cerebral palsy often occurs concurrently with other comorbidity, including "disturbances of sensation, perception, cognition, communication and behavior … epilepsy and … secondary skeletal problems."(47) In the ELGAN study of preterm children born <28 weeks gestation, 11.4% of infants met the criteria for CP, and of those, infants with quadriplegic CP were more than 9 times more likely to be more severely impaired (measured by Bayley-II scores <70, microcephaly, a positive screen for autism, or GMFCS score ≥ 2) than those with diplegic CP.(50) Cerebral palsy, therefore, can be

present with other developmental impairments in the preterm population, and the severity and type of motor deficits associated with cerebral palsy can have a crossover impact on other domains of functional neurodevelopmental ability.

Those who are clinically confirmed to have some form of cerebral palsy are evaluated at age 18-21 months with the Gross Motor Function Classification System (GMFCS) to determine the severity of CP. The GMFCS is a scoring system that attempts to objectively categorize infants with CP in terms of lower limb impairment severity, rather than the specific type of motor disorder. Infants are given a score from 1-5, with higher numbers indicating more severe motor impairment.(51) Descriptions of each classification level can be found in the Appendix (Table 15). The incidence of severe cerebral palsy can range from 5.7% to 2.2% to 1.3% depending upon how severe CP is defined – CP with a GMFCS score ≥2, ≥3, or ≥4, respectively.(42,43,52)

1.4.3 Hearing and vision impairment

Other forms of impairment associated with preterm birth include hearing and vision problems. Retinopathy of prematurity (ROP), for example, is a vision disorder that occurs in VPT survivors and is caused by improper retinal development. In its most severe form, retinal detachment causes blindness.(53) The incidence of bilateral blindness among VPT infants ranges from 0.6% to 1.6% based on studies which use bilateral blindness as their visual outcome, though some give specific acuity cutoffs such as 6/60 in the better eye.(43,54,55) A retinal examination is performed at 6 weeks of age and follow-up is continued

until the retinal vasculature has matured and the risk of progression has passed. Follow-up visits are recommended for the first year after birth. Infants are then categorized as functionally normal, or as having unilateral or bilateral impairment at the two-year follow-up based upon physical exam, ophthalmology reports or parental report.

Preterm infants are at a tenfold increased risk for hearing impairment.(56) Routine clinical care encompasses a newborn hearing screening and an audiology assessment typically at 8-9 months of age. Hearing loss can occur due to a number of reasons, including sensorineural or conductive impairment and auditory neuropathy.(57) The level of hearing loss can be described in terms of audiogram results, a test that measures ability to hear varying frequencies and intensity of sound. For instance, a child may be classified as having mild hearing loss if they are only able to hear and respond to sounds when they reach an intensity level of 26-40 dB. A child with severe hearing loss cannot hear sounds below 71 dB.(57) Auditory impairment may necessitate hearing aids or cochlear implants to improve hearing. However, these tools can sometimes fail to work in children with profound hearing loss (>90 dB).(35,42)

In reference to the level of hearing impairment, there is variation in the literature about what constitutes severe impairment. Some studies suggest that the need for cochlear implants or hearing aids is evidence of severe impairment (54,55), while other studies only categorize those with profound hearing loss (>90 dB) as a severe hearing impairment.(42,52,58) Studies with incongruent

definitions give estimates that vary from incidence rates of 2.5% among very preterm infants under the broader definition, to 1.1% with more stringent criteria. Others studies do not include hearing loss in their composite measure of severe neurodevelopmental impairment.(59,31,46)

1.4.4 Follow-up at school age, adolescence, and into adulthood

Severe neurodevelopmental impairment that is evident at the age of 18-24 months corrected postmenstrual age may not resolve later in life. Many studies have investigated the long-term consequences of very preterm birth by prospectively following-up VPT survivors to school age (60–64) and beyond. Indeed, impairments can persist with varying severity through adolescence (65,66), and adulthood.(67–70) The adverse effects of very preterm birth can therefore affect an individual for the rest of their life, and this presents children and families with unique challenges to overcome well after discharge from the neonatal intensive care unit.

1.5 Severe neurodevelopmental impairment (SNI)

1.5.1 Classification as "severely impaired" using a composite outcome

As mentioned above, very preterm infants are at increased risk for a variety of impairments in different domains of health. The gradation of severity of the impairment within any domain can vary widely, and this affects the prognosis for the infant. Since there are a large variety of possible types and levels of impairment, studying the functional outcomes of VPT infants can be challenging

without classification of children into different categories of impairment severity. Clinicians and researchers alike use cut-off criteria to classify a child as developmentally normal, as having moderate neurodevelopmental impairment (MNI) or as having severe neurodevelopmental impairment (SNI).(14,21,31,42,43,46,54,55,58,59) Follow-up assessments are typically used to identify impairments in five distinct domains: cognition, language, motor, hearing and vision. Classifying infants in this way allows research to have well defined categorical outcomes, making the results of analyses easier to interpret and allowing identification of concrete goals of care.(39) For instance, a hospital may hope to reduce the number of infants with SNI by 10% over a five-year period by implementing a new protocol. Without a simple categorization system for neurodevelopmental impairment, evaluation of success in meeting such targets becomes more challenging.

Furthermore, physicians use impairment rates from follow-up research studies to counsel parents anticipating the birth of a very preterm infant, or shortly after birth. If an infant seems destined to die or survive with severe neurodevelopmental impairment, a clinician will present this information to parents in order to help with the difficult decision to either continue care or withdraw life-saving medical support from their very sick child.(13,71–74) Should a family choose to continue high intensity care of the VPT infant (as opposed to providing palliative care) despite a high probability of SNI, the type and severity of the SNI will indicate what rehabilitation therapies and intervention strategies may be required to improve long-term outcomes and maximize the health and

wellbeing of the child. Predictions about the likelihood of impairment can be made based upon risk factors, and at 18-24 months of age, the earliest time point for reliable neurodevelopmental follow-up assessment, parents can begin to get a more definitive answer with regard to neurodevelopmental status.(41) For these reasons, identifying children who are at high risk of severe neurodevelopmental impairment is critical for informed decision-making and early intervention.

1.5.2 Defining Severe Neurodevelopmental Impairment

Improving the accuracy of quantifying SNI incidence rates in hospitals and research networks requires clarity with regard to the definitional criteria for severe neurodevelopmental impairment. Identifying rates of SNI in very preterm survivors and analyzing outcome variability across hospital sites is difficult when the SNI is not consistently defined. Currently, there are no standard criteria for categorizing infants as impaired in long-term follow-up studies of preterm survivors.

Research into the neurodevelopmental outcomes of preterm survivors often requires dichotomizing or categorizing continuous scores obtained from scales that measure neurodevelopmental function. Composite measures that classify infants as having severe neurodevelopmental impairment lump together children with a wide spectrum of impairments.(54,58,75–77) For example, a child who has a cochlear implant for an isolated hearing impairment may be classified as having a SNI, as would a child with multiple cognitive, motor and sensory

impairments. These children are clearly not equivalent in terms of the severity of their disability. Furthermore, assessment of one domain of neurodevelopment can be hindered by impairments in other domains. For example, the Bayley Scales of Infant and Toddler Development 3rd edition (Bayley-III) is standardized for children with normal sensory function. Assessing motor performance in children who are blind is difficult due to the inability to see obstacles while walking and objects used to test fine motor skills. Similarly, hearing problems are often associated with impaired language development.

It is important to note that there are legitimate reasons to use different definitions. Some study groups may only be interested in the most severe of the impaired infants as their focus is on risk of death and dramatically reduced quality of life for VPT infants. In contrast, other groups may cast a broader net with a more inclusive definition of SNI to capture VPT infants who may survive the neonatal period but face life-long challenges that necessitate more involved patient care. Nevertheless, there is a need for clarity in scientific communication with regard to the types, number and severity of impairments affecting preterm children. This will allow appropriate interpretation of results and a universal approach to study and managing children with adverse neurodevelopmental outcomes.

1.5.3 Impact of inconsistent SNI definition

Criteria for what should be considered severe impairment, as opposed to mild or moderate impairment, are not consistent across the literature. For

instance, some study groups consider a Bayley-III score that is two standard deviations below the mean to indicate SNI, whereas other groups only consider scores below three standard deviations to be severe impairment.(42,46,55,58) Varying the cutoff-criteria and grouping of impairments when deciding upon definitions of SNI may have a significant effect on the numbers and rates of SNI. Nevertheless, the effect of SNI definition on incidence rates of SNI is understudied, and it is unclear how incidence rates of SNI are affected by the definition of SNI. Furthermore, though many studies report associations between risk factors (Figure 1) and neurodevelopmental impairment, it is unclear whether the associations change or remain unchanged when the definition of SNI varies (i.e., more inclusive vs. exclusive criteria for SNI).

If incidence rates and strength of association with risk factors vary depending upon how SNI is defined, the true epidemiological impact of neurodevelopmental impairment in very preterm survivors becomes difficult to discern. Critical evaluation through benchmarking at a program, regional, or national level often requires that comparison be made with other programs, regions or countries with a similar population. Comparisons of SNI rates may not be possible when the definition of SNI varies from one jurisdiction to the next. Furthermore, when designing research studies, in particular randomized control trials, the incidence rate of the outcome of interest determines the sample size necessary to achieve appropriate statistical power. It is therefore critical for clinicians and epidemiologists to arrive at an acceptable and standardized definition of SNI.

Finally, since the incidence of SNI is dependent on how it is defined, the clinical implications of the SNI definition will vary depending on the SNI definition. Clinicians and families need to understand that an infant classified as having SNI in a country or region that uses a broad definition of SNI may have SNI of a lesser severity than an infant in a country or region with a very narrow definition that only captures the most impaired infants. In this respect, the definition of neurodevelopmental impairment is vitally important for decision making and counseling, particularly in instances where life-support may be withdrawn from a critically ill VPT infant. The current literature is unclear to what extent varying definitions of SNI impact incidence rates and how the varying definitions alter associations between putative risk factors and SNI. These issues are important for families and the international community invested in improving the care and long-term outcomes of very preterm infants.

2. Research Question

What is the impact of the definition of severe neurodevelopmental impairment on incidence rates and strength of association between risk factors and neurodevelopmental outcome at 18-21 months corrected age among very preterm survivors born at less than 29 weeks gestation in Canada?

2.1 Objectives

1) To identify definitions of severe neurodevelopmental impairment (SNI) at 18 to 21 months corrected gestational age used by national networks reporting on health outcomes of very preterm babies in the last ten years.
2) To examine the effect of varying definition on the incidence rates of severe neurodevelopmental impairment.
3) To examine the effect of the severe neurodevelopmental impairment definition on the strength of association between known risk factors and severe neurodevelopmental impairment.

2.2 Hypotheses

1) There will be a significant difference in incidence rates of severe neurodevelopmental impairment dependent upon the definition of severe neurodevelopmental impairment.
2) The strength of association between risk factors and severe neurodevelopmental impairment will vary depending on the definition of severe neurodevelopmental impairment that is used.

3. Methods

3.1. Severe Neurodevelopmental Impairment definitions

A literature and internet search of major neonatal research networks that follow up very preterm infants was performed to identify definitions of severe neurodevelopmental impairment (SNI). Networks were identified by searching for the key words “neonatal network” and “neonatal follow-up” in a Google search, as well as individual entry or combinations of the terms “neonatal,” “premature,” “prematurity,” “preterm,” “disability,” “impairment,” and “follow-up” in a PubMed search. The literature search was limited to studies or reports published in the past 10 years to target SNI definitions that are currently used in clinical practice. Neonatal networks were included if they met the following criteria: 1) focus on very preterm infants; 2) systematic follow up of very preterm infants; 3) the network had more than one site and more than one investigator involved; 4) the network published more than one research protocol in scientific literature in English; and 5) included follow-up information on neurodevelopmental outcome (including moderate or severe neurodevelopmental impairment components) at the 18-24* month assessment (Appendix: Table 16). Regional networks that were part of a large national or international network (for example, regional networks in London or Manchester, England, that are part of the larger British Association of Perinatal Medicine network), were not included in the review since their definitions for neurodevelopmental outcome followed the same guidelines and criteria as the larger network.

Definitions of severe neurodevelopmental impairment used by neonatal networks that met our inclusion criteria were abstracted, including specific criteria for each domain of developmental assessment, such as the Bayley-III and GMFCS, and functional criteria for hearing and vision impairment. All definitions of SNI published in the scientific literature were abstracted, including modifications that may have occurred within individual networks over time.

3.2 Study population and data source

The main data source for this study was the Canadian Neonatal Network (CNN) database and the Canadian Neonatal Follow-Up Network (CNFUN) database that included information on infants admitted to participating Neonatal Intensive Care Units (NICU) across Canada and the results of neurodevelopmental assessment of children at 18-21 months of age (corrected for gestational age at birth).(55,78)

3.2.1 Canadian Neonatal Network (CNN) and Canadian Neonatal Follow-Up Network (CNFUN)

The Canadian Neonatal Network (CNN^{TM}) includes a group of Canadian neonatologists and researchers that collaborate on advancing clinical care for critically ill neonates. Dr. Shoo Lee founded the network in 1995 in order to collect national data on admissions to tertiary level neonatal intensive care units in order to study variations in practices, resource use, and outcomes. In 2014, CNN included 31 hospitals and 17 universities across Canada that collected information on infants admitted to participating neonatal intensive care units (NICU). Only

infants admitted for more than 24 hours have detailed data included in the CNN database.

The Canadian Neonatal Follow-Up Network (CNFUN) functions in liaison with the CNN.(79) CNFUN sites recruit infants born at less than 29 weeks gestation who were admitted to a CNN-participating NICU and evaluate neurodevelopmental progress of these children at 18-21 months of age using standardized age-appropriate developmental assessment tools. Data collected by CNFUN include information about caregiver socio-demographic characteristics, medical history regarding visual and hearing problems and health resource utilization since NICU discharge, and results of the neurodevelopmental follow-up assessment. The CNN and CNFUN databases were linked using a unique identifier in a project entitled “The CIHR Team in Maternal Infant Care” (MiCare) to facilitate research on the developmental trajectories of these very preterm born infants. All CNN and CNFUN sites obtained ethics approval for the MiCare study from their local institutional ethics review board. In Vancouver, ethics approval was obtained from the Research Ethics Board of University of British Columbia and the Children’s and Women’s Hospital and Health Centre of British Columbia (H08-02153 and H03-70448).

The linked database is housed at the MiCare coordinating centre at Mount Sinai Hospital/University of Toronto in Toronto, Ontario. No chart review was performed in this study, and data queries for statistical analyses were submitted to the MiCare site and performed by personnel at the coordinating site. Only aggregated data and results from the statistical analyses were made available to

researchers outside the MiCare research site. The following information on clinical characteristics during pregnancy, delivery, and infant's NICU stay were used in analyses (Appendix: Table 17):

- Socio-demographic data: primary caregiver's age, education level, employment status, and ethnic group, and single parenthood status.
- Pregnancy data: maternal drug use, cigarette use, parity, diabetes, pre-eclampsia, antenatal corticosteroid treatment (for fetal maturation), and multi-fetal gestation.
- Delivery data: month and year of birth, location of delivery, mode of delivery, presentation, outborn status (i.e., infant was transferred to NICU from another hospital), delayed cord clamping, gestational age at birth, birth weight, small for gestational age (SGA), sex, inotrope exposure (e.g. dobutamine, epinephrine, milrinone), severity of illness (SNAP-II) score, resuscitation after delivery, 5 minute Apgar score, and duration of oxygen support.
- NICU course data: medical complications, intraventricular hemorrhage (IVH, all grades and severe grades III-IV), sepsis (positive blood or cerebrospinal fluid culture), bronchopulmonary dysplasia (BPD, oxygen requirement at 36 weeks corrected age), retinopathy of prematurity (ROP), necrotizing enterocolitis (NEC) with a Bell's grade ≥ 2, and patent ductus arteriosus (PDA).

- Assessment at 18-21 months corrected age (outcome data): cerebral palsy and GMFCS severity rating, Bayley-III cognitive, motor and language composite scores, hearing or visual impairment.

3.2.2 Study population

This study included infants born at 22-28 weeks gestation between April 1, 2009 and Sept 30, 2011 admitted to a participating neonatal intensive care unit (NICU), who were followed-up by the CNFUN network and assessed for neurodevelopmental impairment. We excluded infants who were stillborn, moribund, admitted to NICU after 24 hours following birth, and infants born with major congenital anomalies (Figure 1). The target age of follow-up was 18-21 months, but infants were not excluded if they were seen after 21 months.*

*NOTE: The most common follow-up age for VPT infants in the literature is 18-24 months of age – when not referring specifically to CNFUN data, "18-24 months" is used to refer to the time of follow-up assessment for VPT survivors more generally.

3.3. Analysis

3.3.1. Severe neurodevelopmental impairment definitions (Objective 1)

We compiled the literature and web search results and assessed them using the inclusion/exclusion criteria described above. Networks that were identified are presented in Table 1. Reasons for exclusion are described if inclusion criteria were not met.

3.3.2. Variation in incidence of severe neurodevelopmental impairment by definition (Objective 2)

3.3.2.1. Descriptive Statistics

Descriptive analyses were performed to examine demographic and clinical characteristics of mothers and their infants. Frequency tables were used for categorical data, while mean, standard deviation, median, interquartile range, minimum and maximum were used to describe continuous variables such as birth weight (Tables 2 and 3).

3.3.2.2. Incidence rates of severe neurodevelopmental impairment by domain

We described results of 18-21 month assessment of neurodevelopmental impairment among CNFUN infants, with a focus on each domain or component of SNI. Analyses of individual domains of severe neurodevelopmental impairment incidence were calculated using the current criteria for SNI used by CNFUN (Table 4). The number of domains impaired was also tabulated for each child and cross-tabulated by affected domain (Table 5).

3.3.2.3. Incidence rates of severe neurodevelopmental impairment by varying definition

The overall incidence of severe neurodevelopmental impairment in the CNFUN dataset was calculated using the criteria identified in different definitions for SNI. Crude rate ratios (RR) and 95% confidence intervals (CI) were

calculated to compare incidence rates of SNI defined by CNFUN versus other networks (Table 6).

3.3.3 Strength of association between maternal and infant risk factors and severe neurodevelopmental impairment (Objective 3)

3.3.3.1. Descriptive analysis and bi-variate analyses

In bi-variate analyses, potential risk factors were examined for their association with SNI for the definitions that had the lowest and highest incidence rates (less inclusive SNI and more inclusive SNI definitions, respectively). Chi-square or Fisher's exact tests were used to assess statistical significance of the association for categorical variables, and the t-test or Wilcoxon sum-rank test was used for continuous variables (Table 7).

3.3.3.2 Multivariable analyses

Logistic regression models were used to examine independent associations between known risk factors and SNI using the same two definitions of SNI as were used in the bi-variable analyses (i.e., the least inclusive SNI and most inclusive SNI definition). Individual risk factors were entered into the logistic regression model in several stages. In the first stage of the regression analysis, risk factors that were most distal to the outcome were modeled (i.e., antenatal risk factors), while each subsequent stage included progressively more proximal factors (e.g., delivery characteristics, neonatal morbidity, etc.). At each stage, an alpha level of $p<0.1$ was used to retain variables that were significantly associated with SNI. Variables with results above this alpha level threshold were

deemed to be not associated with SNI and were removed from the model. Variables that met the inclusion criteria after the first stage were retained in the model for the second stage regression analysis. This process was repeated until the final stage was completed. The variables included each stage were as follows:

- Stage 1: Prenatal characteristics
    - Variables: Ethnicity, employment, education, single parenthood, fertility treatment, antenatal steroid exposure, drug use during pregnancy
    - Inclusion criteria = $p<0.1$
- Stage 2: Birth risk factors
    - Variables: Apgar at 5 minutes <7, gestational age (continuous), small for gestational age (<10^{th} percentile), infant's sex, outborn status
    - Inclusion criteria = $p<0.1$
- Stage 3: Illness severity
    - Variables: SNAP-II score <20, late-onset sepsis, bronchopulmonary dysplasia, retinopathy of prematurity, necrotizing entercolitis
    - Inclusion criteria = $p<0.1$
- Stage 4: Brain injury
    - Variables: Grade 3 or 4 intraventricular hemorrhage (IVH) or periventricular leukomalacia (PVL)
    - Inclusion criteria = $p<0.1$

The logistic regression model that included all significant risk factors after stage 4 was considered the final parsimonious model. For this final model we used the standard p-value cut-off <0.05 to identify statistically significant associations (Tables 8 and 9).

Adjusted odds ratios (AOR) and 95% CIs were calculated to quantify the association between each risk factor and severe neurodevelopmental impairment. Adjusted odds ratios from models with SNI based on the most inclusive definition of SNI and the least inclusive definition of SNI were compared to determine whether or not there were differences in the strength of the association depending on SNI definition.

For each SNI definition, an extended regression model was created that included all variables selected in the final models (for both definitions of SNI). In other words, all factors that were found significantly associated with SNI using the more inclusive definition of SNI were employed in the regression model with the less inclusive SNI definition as the outcome, and vice-versa (Table 10). Thus all factors significantly associated with either the less or the more inclusive SNI definition were compared.

Hosmer-Lemeshow test and c-statistics were used to evaluate goodness-of-fit of the logistic regression models. The Hosmer-Lemeshow test assesses goodness-of-fit by dividing observations into deciles based on estimated probabilities of the outcome, and then providing a chi-square statistic from observed and estimated frequencies. A Hosmer-Lemeshow p-value <0.05 indicates statistically significant differences between predicted and observed

outcomes (i.e., poor goodness-of-fit). C-statistics show whether predicting the outcome using the regression model is better than chance (a value of 0.5 indicates no better than chance prediction, while 1 indicates a perfect prediction of the outcome). C-statistic between 0.7 and 0.8 indicates good classification accuracy, and c-statistics over 0.8 indicate a very good classification accuracy of the model.(80)

3.3.4 Additional analyses

3.3.4.1 Site differences

Rates of SNI were compared between participating CNN/CNFUN sites using both SNI definitions (most inclusive SNI and least inclusive SNI definitions) (Table 11). CNN/CNFUN sites were ranked by decreasing SNI incidence under both SNI definitions. The number and proportion of sites that changed their rank as a result of a changed SNI definition were described (Table 12).

3.3.4.2 Missing values

Missing data were described by calculating the proportion of subjects with missing data for each SNI domain (Table 13), as well as the proportion of infants with missing data in multiple domains (Table 14). For instance, if an infant was not screened for vision and hearing problems but otherwise had complete data, then the infant would have missing data in two domains. No attempt was made to impute missing values, and all infants who met inclusion criteria for the study were included, irrespective of incomplete/complete information on all SNI

domains. However, in domain specific analyses, infants with missing data in that particular domain were not included in the denominator for calculation of incidence rates.

4. Results

4.1 Severe Neurodevelopmental Impairment Definitions (Objective 1)

4.1.1 Identification of Eligible Neonatal Research Networks

The literature and web search for neonatal networks that may have a definition suitable for the study identified 29 different national or regional neonatal networks. Of these, only eight met the inclusion criteria. Seventeen neonatal networks were excluded for the following reasons: did not carry out long-term follow-up (n=9) or assessments were at an age outside 18-24 months corrected age (n=4) (Table 1). For instance, the Japanese Neonatal Research Network followed-up very preterm survivors at the age of 3 years.(81,82) Furthermore, some research groups were in the process of obtaining 2-year follow-up information, but had not published results to date (n=2), while others did follow-up at 2 years of age but used developmental tests different from those used by CNFUN (n=2), such as the Ages and Stages Questionnaire or the Brunez-Lézine Test (66), and thus were not directly comparable with CNFUN data.

The eight research networks that met the inclusion criteria for the study were: the Australian and New Zealand Neonatal Network (ANZNN), the British Association of Perinatal Medicine (BAPM), the Canadian Neonatal Follow-Up Network (CNFUN), the Extremely Preterm Study Group (EXPRESS), the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network (NICHD), the Swiss Neonatal Network (SNN), the Vermont Oxford Network (VON), and the Victorian Infant Collaborative Study (VICS).

Table 1: Neonatal networks (NN) identified in literature review and reasons for exclusion if inclusion criteria not met

Networks that met study criteria	
Australia New Zealand Neonatal Network (ANZNN) British Association of Perinatal Medicine (BAPM) Canadian Neonatal Follow-Up Network (CNFUN) Eunice Kennedy Shiver National Institute of Child Health and Human Development Neonatal Research Network (NICHD) Extremely Preterm Study Group (EXPRESS) Swiss Society of Neonatology (SwissNeoNet) Vermont Oxford Network (VON) Victorian Infant Collaborative (VICS)	
Networks that did not meet study criteria	**Reason for Exclusion**
Egyptian NN Gaza NN Italian NN Israeli NN Korean NN NEOCOSUR (South America) Saudi NN SEN1500 (Spain) SIBEN (Ibero-American NN) Norwegian Extreme Prematurity Study Group	No long-term neurodevelopmental follow-up
Japan Neonatal Research Network (NRNJ) National Collaborative Perinatal Neonatal Network (Lebanon) CMF NNN (Chinese National NN) Swedish National Quality Register	Neurodevelopmental follow-up not between 18-24 months corrected gestational age
German Neonatal Network Scottish Neonatal Network	Neurodevelopmental follow-up ongoing but not yet published
EPIPAGE (France) French Regional Loire Infant Follow-Up Team (LIFT)	Non-comparable developmental tests (e.g. Brunez-Lézine)

4.1.2 Identification of severe neurodevelopmental impairment definitions

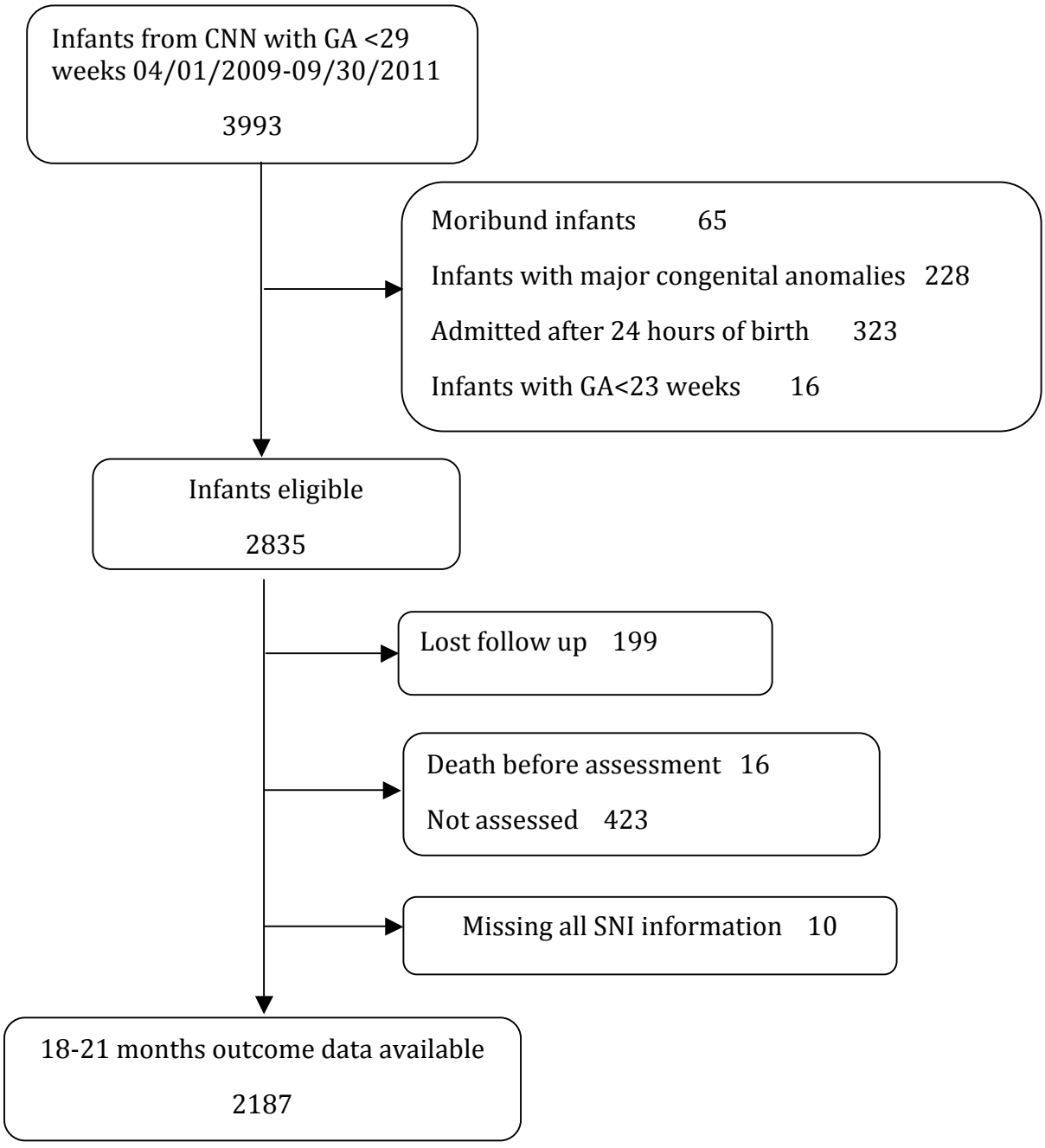
Among the eight eligible neonatal follow-up networks, seven distinct definitions for severe neurodevelopmental impairment were identified. Definitions included criteria pertaining to six different domains: 1) cerebral palsy severity (using the Gross Motor Function Classification System (GMFCS)), 2) motor, 3) language, and 4) cognitive Bayley-III composite scores, and 5) visual and 6) hearing impairment. Each network had specific criteria for impairment in each functional domain that was deemed severe. Differences in the definition of SNI were relatively small with regard to some domains, while other domains showed larger differences. For instance, GMFCS scores of 3-5 were considered severe in more inclusive definitions, compared with a score of 4-5 in more stringent definitions. In contrast, definitional differences in Bayley-III scores were larger, e.g., using 2 vs. 3 standard deviations below the mean as the threshold for severe impairment.

4.2 Study population

Between April 1, 2009 and September 30, 2011 there were 3993 infants born at <29 weeks gestation who were admitted to a CNN NICU. Sixty five infants were moribund, 228 infants had major congenital anomalies, 323 were admitted >24 hours after birth, 16 infants were born <23 weeks GA, and 526 died before discharge, totaling 1158 infants who were excluded from the cohort. Of the 2835 remaining eligible infants, 199 were lost to follow-up, 16 died before assessment, 423 were eligible but not assessed, and 10 were assessed but had

missing data for all relevant SNI domains (Figure 1). The final cohort consisted of 2187 infants who were followed-up by CNFUN (55%).

Figure 1: Flow diagram of CNN/CNFUN infants who met inclusion criteria



4.3 Analyses

4.3.1 Descriptive statistics of cohort

A summary of the descriptive statistics is shown in Tables 2 and 3. Descriptive characteristics of the study population were summarized under four categories: socio-demographic factors, pregnancy, delivery and birth factors, and severe morbidity diagnosed during the infant's stay in the NICU.

Table 2: Descriptive statistics for potential risk factors of severe neurodevelopmental impairment (nominal)

Risk factor	N	%	Min	Max	Mean	SD	Median	IQR
Pregnancy								
Maternal age (years)	2089	95.5	15	56	30.9	5.9	31	27-35
Delivery and Birth								
GA (weeks)	2187	100	23	28	26.4	1.4	27	25-28
Birth weight (g)	2186	100	420	1805	934	222	920	770-1090
NICU								
SNAP-II score	2175	99.5	0	75	15.0	11.5	14	9-21

Table 3: Descriptive statistics for potential risk factors of SNI (categorical)

Risk factor	N (n=2187)	%	Missing	%
Socio-demographic				
Ethnicity:			0	0
First Nations	77	3.5		
East Asian	60	2.7		
South Asian	144	6.6		
White	1260	57.6		
Other/unknown	646	29.5		
Education:			169	7.7
Less than high school	220	10.9		
Completed high school	696	34.5		
Post-secondary	1102	54.6		
Employment status:			0	0
Unemployed	102	4.7		
Student or homemaker	654	29.9		
Employed part-/full-time	1215	55.6		
Other/unknown	216	9.9		
Lone parent	151	7.0	32	1.5
Pregnancy				
Maternal age			98	4.5
<25	308	14.7		
25-34	1280	61.3		
35+	501	24.0		
Parity (primipara)	790	36.3	12	0.5
Diabetes	171	8.1	81	3.7
Preeclampsia	370	17.2	37	1.7
Antenatal steroids	1963	91.4	43	2.0
Drug use during pregnancy	72	3.3	0	0
Cigarette use in pregnancy	259	11.9	0	0
Twins/multiple gestation	630	28.8	0	0

Risk factor	N (n=2187)	%	Missing	%
Delivery and Birth				
Apgar score 5min <7	766	35.3	17	0.8
SGA	159	7.3	4	0.2
Sex (male)	1159	53.0	2	0.1
Clinical chorioamnionitis	466	25.3	347	15.9
PROM >24hours	469	21.7	30	1.4
Mode of delivery:			6	0.3
Vaginal	911	41.8		
Cesarean	1270	58.2		
Presentation:			5	0.2
Vertex	1230	56.4		
Breech	696	31.9		
Transverse	122	5.6		
Other	134	6.1		
Outborn	216	9.9	2	0.1
NICU				
SNAP-II score>20	590	27.1	12	0.5
IVH- all grades	798	36.9	23	1.1
IVH – severe (grade 3-4)	224	10.5	54	2.5
Sepsis (positive blood/CSF):			0	0
Early onset	37	1.7		
Late onset	617	28.2		
BPD at 36 weeks	987	45.2	1	0
ROP – all	1005	57.4	437	20.0
ROP – severe (≥ stage 3)	259	15.2	481	22.0
NEC – Bell's ≥ grade 2	152	7.0	7	0.3
PDA – treated	1004	46.4	24	1.1

4.3.1.1 Socio-demographic factors

Socio-demographic factors included in analyses were primary caregiver's ethnicity, education level, employment status, and single parent status. In the CNFUN cohort, a majority of primary caregivers were white (57.6%), highly educated (54.6% post-secondary), and employed part- or full-time (55.6%). Other ethnicities included South Asian (6.6%), First Nations (3.5%), and East Asian (2.7%). Eleven percent did not complete high school, and 34.5% completed high school but did not continue on to post-secondary education. Those who were not employed were primarily students or homemakers (29.9%). There were 152 caregivers who identified as single parents (7.0%). Of note, there was a sizeable proportion of caregivers whose ethnicity and employment status were labeled "other" and "unknown" (29.6% and 9.9%, respectively).

4.3.1.2 Pregnancy characteristics

The average maternal age was 30.9 years (SD 5.9 years) and ranged from 15-56 years. When stratified by maternal age categories, 14.7% of women were <25 years old, 61.3% were between 25-34 years of age, and 24.0% were 35 years or older. While most infants were singletons (71.2%) twins or higher order multiple gestation pregnancies accounted for 28.8%. More than one-third of women were primiparous (36.3%). One hundred and seventy-one mothers had chronic or gestational diabetes (8.1%), and 17.2% experienced preeclampsia. The vast majority of fetuses in the cohort were treated with antenatal steroids

(91.4%); 11.9% of infants were born to mothers who reported smoking cigarettes during pregnancy, and 3.3% were exposed to illicit drugs while in the womb.

4.3.1.3 Delivery and Birth characteristics

One in ten deliveries were “outborn”, meaning infants needed to be transported to a tertiary level NICU. We included only those admitted within 24 hours of delivery. A majority of infants were born by caesarean delivery (58.2%) and vertex presentation at birth was most common (56.4%), as compared with breech (31.9%), transverse (5.6%) or other fetal presentations (6.1%). Approximately a quarter of mothers (25.3%) had a delivery complicated by clinical chorioamnionitis; 21.7% of mothers had prolonged rupture of membranes (PROM) lasting for more than 24 hours. Mean gestational age (GA) at birth was 26.4 weeks (SD 1.4 weeks, range 23-28 weeks), and mean birth weight was 934 grams (SD 222 g, range 420-1805 g). Overall, 53.0% of infants were male, 159 infants (7.3%) were small for gestational age (SGA), and 35.3% of infants had a 5-minute Apgar score <7.

4.3.1.4 Severe morbidity during NICU stay

As expected in the very preterm population, most surviving infants experienced one or more morbidities while in the NICU. Nearly six hundred infants (27.1%) had a high initial Score for Neonatal Acute Physiology at birth (SNAP-II score >20); 46.4% were medically or surgically treated for a patent ductus arteriosus (PDA) 29.9% were diagnosed with sepsis (1.7% had early

onset, and 28.2% had late onset sepsis), and 7.0% had necrotizing entercolitis. Almost half of all VPT survivors (45.2%) required oxygen or respiratory support at 36 weeks post-menstrual age (i.e, had BPD). Severe retinopathy of prematurity (grade 3 or more) was diagnosed in 15.3% infants, and intraventricular hemorrhage was present in 798 infants, 224 of whom had severe brain injury – grade 3 or 4 IVH (10.5%).

4.3.2 Incidence of severe neurodevelopmental impairment (Objective 2)

4.3.2.1 Incidence rates of severe neurodevelopmental impairment by domain of impairment

Exploratory analyses of individual domains of severe neurodevelopmental impairment incidence were calculated using the current criteria for SNI used by CNFUN (Table 4). Forty-eight (2.2%) infants were diagnosed with severe cerebral palsy (GMFCS 3-5). Using a threshold of <2SD below the mean (score <70) for Bayley-III composite scores, 93 (4.3%) surviving infants had severe motor impairment, 59 (2.7%) had severe cognitive impairment, and 188 (8.6%) had severe language impairment. There were 56 (2.6%) infants with hearing aids or cochlear implants and 32 (1.5%) who were bilaterally blind. Infants with impairment in one domain were often impaired in other domains. The frequency of overlapping impairments is displayed in Table 4.

Table 4: Incidence of severe neurodevelopmental impairment (SNI) by domain using CNFUN definition (≥1 of: GMFCS 3-5; Bayley-III <70 for motor, cognitive, or language composite score; hearing aids or cochlear implant, bilateral blindness)

	Severe Impairment (n/denominator)					
SNI Domain	CP (GMFCS)	Motor (Bayley)	Cognitive (Bayley)	Language (Bayley)	Hearing (Aids)	Vision (Blind)
CP (GMFCS)		13/48 [33]	13/48 [24]	14/48 [26]	12/48 [1]	12/48 [5]
Motor (Bayley)	13/93 [5]		18/93 [1]	43/93 [2]	6/93 [5]	5/93 [16]
Cognitive (Bayley)	13/59 [2]	18/59 [24]		42/59 [2]	5/59 [3]	10/59 [6]
Language (Bayley)	14/188 [8]	43/188 [29]	42/188 [3]		10/188 [10]	11/188 [23]
Hearing	12/56 [2]	6/56 [17]	5/56 [13]	10/56 [18]		6/56 [5]
Vision	12/32 [2]	5/32 [19]	10/32 [10]	11/32 [10]	6/32 [0]	

Each cell shows the number of cases and the number of children in the specified domain of impairment. The numbers in parenthesis refer to the number of children with missing information. For example, of the 32 infants were bilaterally blind, 12 were also diagnosed with severe grade cerebral palsy and 2 infants had missing CP data, as seen in the last row of the second column.

The number of domains impaired was also tabulated for each child and cross-tabulated by affected domain. Among the 48 infants with severe cerebral palsy, 12 had no other impairments, 17 had one other impairment, 11 had 2 other impairments, etc. (Table 5). These results show the number of infants with severe impairments in multiple health domains and who may consequently have less autonomy due to restricted functional ability. It is evident that many infants were adversely affected in more than one domain: no other impairments were

found in only 25% of infants with CP, 37% with motor impairment, 12% with cognitive impairment, 59% with language impairment, 54% with requirement of cochlear implants or hearing aids, and 25% with bilateral blindness.

Table 5: Frequency of multiple domains of impairment using CNFUN definition of severe neurodevelopmental impairment (SNI)

SNI Domain	Severe Impairment (n/denominator)					
	No other	1 other	2 other	3 other	4 other	5 other
CP (GMFCS)	12/48	17/48	11/48	7/48	1/48	0
Motor (Bayley)	34/93	40/93	13/93	5/93	1/93	0
Cognitive (Bayley)	7/59	26/59	17/59	8/59	1/59	0
Language (Bayley)	110/188	47/188	21/188	9/188	1/188	0
Hearing	30/56	16/56	8/56	1/56	1/56	0
Vision	8/32	10/32	8/32	6/32	0	0

4.3.2.2 Incidence rates of severe neurodevelopmental impairment by definition

SNI incidence rates ranged from 3.5% to 14.9% depending on the definition of SNI used [Rate Ratio = 4.29; 95% CI: 3.37 5.47)] (Table 6). In other words, the most inclusive definition yielded an incidence rate that was more than four times greater than incidence rate obtained by the least inclusive definition when applied to the same population.

Table 6: Incidence of severe neurodevelopmental impairment by definition (N=2187*)*

Definition	Motor CP (GMFCS)	Motor Score (Bayley-III)	Language Score (Bayley-III)	Cognitive Score (Bayley-III)	Hearing	Vision	SNI No.	SNI (%)	Rate Ratio	95% Confidence Interval
CNFUN & VON	3-5	<70	<70	<70	Hearing aid or cochlear implant	Bilateral blindness	326	14.9	4.29	3.37, 5.47
VICS	4-5	—	<64	<68	—	Bilateral blindness	183	8.4	2.40	1.86, 3.13
NICHD	4-5	<70	—	<70	Profound hearing loss (>90dBHL)	Bilateral blindness	170	7.8	2.24	1.72, 2.91
BAPM & Swiss NN	3-5	—	≤55	<55	Profound hearing loss (>90dBHL)	Bilateral blindness	92	4.2	1.21	0.9, 1.63
ANZNN 2013	4-5	≤55	≤55	≤55	—	Bilateral blindness	89	4.1	1.17	0.87, 1.58
EXPRESS	4-5	<55	<55	<55	Profound hearing loss (>90dBHL)	Bilateral blindness	85	3.9	1.12	0.83, 1.52
ANZNN 2012	4-5	—	<55	<55	—	Bilateral blindness	76	3.5	Ref	Ref

Abbreviations: CNFUN (Canadian Neonatal Follow-Up Network)[55]; VON (Vermont Oxford Network)[84]; VICS (Victorian Infant Collaborative Study)[59]; NICHD (Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network)[42]; BAPM (British Association of Perinatal Medicine)[58]; Swiss NN (Swiss Neonatal Network)[22,85] ; ANZNN (Australia New Zealand Neonatal Network)[31,46] ; EXPRESS (Extreme Preterm Study Group)[86]; GMFCS (Gross Motor Function Classification System)[51]

4.3.3 Strength of association between maternal and infant risk factors and severe neurodevelopmental impairment (Objective 3)

4.3.3.1 Bi-variable analyses

Bi-variable analysis revealed significant associations between several risk factors and severe neurodevelopmental impairment. The results are summarized in Table 7, showing the relationship between risk factors and SNI, using two definitions; SNI with the most inclusive vs. least inclusive definition. SNI, as defined by both the most and least inclusive SNI definitions, was associated with education level, employment status, antenatal steroid exposure, illicit drug use, 5 minute Apgar score, gestational age, birth weight, male sex, SNAP-II score, IVH, late onset sepsis, BPD, ROP, and NEC. However, there were some notable differences across the two definitions in terms of risk factor-SNI associations. For sociodemographic variables, both ethnicity and education level were found to be significantly associated with SNI under the most inclusive definition ($p<0.01$ and $p<0.01$, respectively) but not with SNI as defined by the least inclusive definition ($p=0.28$ and $p=0.10$, respectively). Similarly, maternal age (continuous), primiparity, and treated PDA were significantly associated with SNI as defined by the more inclusive definition ($p=0.03$, $p=0.02$ and $p<0.01$, respectively) but not with SNI under the least inclusive definition of SNI ($p=0.21$, $p=0.08$ and $p=0.73$, respectively). Smoking during pregnancy and antenatal steroid treatment were also associated with the SNI under the most inclusive SNI definition ($p=0.05$, and $p<0.01$, respectively), but not with SNI under the least inclusive definition of SNI ($p=0.15$ and $p=012$, respectively). Outborn status was the only variable

significantly associated with SNI under the least inclusive SNI definition (p=0.03) but not with SNI under the most inclusive SNI definition (p=0.24). Maternal age as a categorical variable, chronic or gestational diabetes, preeclampsia, multiple gestation, SGA, clinical chorioamnionitis, PROM (>24 hours), mode of delivery, delivery presentation, and early onset sepsis were not significantly associated with SNI under either definition of SNI.

Table 7: Bi-variable associations: Association between potential risk factors and severe neurodevelopmental impairment (SNI) for the most and least inclusive definitions (yielding the highest and lowest SNI incidence)

	Severe Neurodevelopmental Impairment			
	Most inclusive definition		Least inclusive definition	
Risk factor	N (%)	p-value	N (%)	p-value
Socio-demographic				
Ethnicity:		<0.01		0.28
First Nations	22 (28.6)		6 (7.8)	
East Asian	6 (10.0)		3 (5.0)	
South Asian	24 (16.7)		4 (2.8)	
White	165 (13.1)		41 (3.3)	
Other/unknown	109 (16.9)		22 (3.4)	
Education:		<0.01		0.10
Less than high school	45 (20.5)		10 (4.6)	
Completed high school	110 (15.8)		29 (4.2)	
Post-secondary	132 (12.0)		28 (2.5)	
Employment status:		<0.01		0.08
Unemployed	24 (23.5)		3 (2.9)	
Student or homemaker	121 (18.5)		32 (4.9)	
Employed part- / full-time	136 (11.2)		32 (2.6)	
Other/unknown	45 (20.8)		9 (4.2)	
Lone parent	33 (21.9)	0.01	10 (6.6)	0.02
Pregnancy				
Maternal age in years (mean, sd)	30.2 (5.9)	0.03	30.0 (6.1)	0.21
<25	59 (19.2)	0.08	13 (4.2)	0.40
25-34	182 (14.2)		47 (3.7)	
35+	72 (14.3)		13 (2.6)	
Parity (primipara)	99 (12.5)	0.02	20 (2.5)	0.08
Diabetes	25 (14.6)	0.97	6 (3.5)	0.89
Preeclampsia	51 (13.8)	0.55	16 (4.3)	0.23
Antenatal steroids	276 (14.1)	<0.01	64 (3.2)	0.12
Drug use during pregnancy	24 (33.3)	<0.01	8 (11.1)	<0.01
Cigarette use in pregnancy	49 (18.9)	0.05	13 (5.0)	0.15
Multiple gestation	102 (16.2)	0.28	21 (3.3)	0.82

Risk Factor	Most inclusive definition N (%)	p-value	Least inclusive definition N (%)	p-value
Delivery and Birth				
Apgar at 5 min <7	148 (19.3)	<0.01	40 (5.2)	<0.01
GA in weeks (mean, SD)	25.8 (1.4)	<0.01	25.6 (1.5)	<0.01
Birth weight in grams (mean, SD)	868 (202)	<0.01	867 (211)	<0.01
SGA	22 (13.8)	0.70	4 (2.5)	0.65
Sex (male)	211 (18.2)	<0.01	50 (4.3)	0.02
Clinical chorioamnionitis	66 (14.2)	0.92	11 (2.4)	0.29
PROM	60 (12.8)	0.15	10 (2.1)	0.08
Mode of delivery:		0.07		0.86
Vaginal	151 (16.6)		31 (3.4)	
Caesarean delivery	175 (13.8)		45 (3.5)	
Presentation:		0.41		0.94
Vertex	172 (14.0)		44 (3.6)	
Breech	117 (16.8)		22 (3.2)	
Transverse	17 (13.9)		5 (4.1)	
Other	20 (14.9)		5 (3.7)	
Outborn	38 (17.6)	0.24	13 (6.0)	0.03
NICU				
SNAP-II score (median, IQR)	16 (9, 25)	<0.01	21 (14, 29)	<0.01
SNAP-II score >20	133 (22.5)	<0.01	39 (6.6)	<0.01
IVH- all grades	189 (23.7)	<0.01	48 (6.0)	<0.01
IVH – severe (grade 3-4)	90 (40.2)	<0.01	32 (14.3)	<0.01
Sepsis – positive blood/CSF				
Early onset	8 (21.6)	0.25	1 (2.7)	0.99
Late onset	128 (20.8)	<0.01	33 (5.4)	<0.01
BPD at 36 weeks	193 (19.6)	<0.01	53 (5.4)	<0.01
ROP – all	207 (20.6)	<0.01	53 (5.3)	<0.01
ROP – severe (≥ stage 3)	73 (28.2)	<0.01	24 (9.3)	<0.01
NEC – Bell's ≥ grade 2	34 (22.4)	<0.01	10 (6.6)	0.03
PDA – medically/surgically treated	191 (19.0)	<0.01	47 (4.7)	0.73

4.3.3.2 Multivariable logistic regression

The association between risk factors and SNI varied depending on the definition used. Maternal ethnicity, employment status, antenatal steroid exposure, and infant gestational age (GA), showed inconsistent associations with SNI, depending on the SNI definition (Table 8 and 9). Because the final models for each definition include different sets of risk factors, the results discussed in this section represent the 'final extended' models (Table 10), which included all variables that showed association with SNI for either definition, and thus allow direct comparison.

Infants born to First Nations mothers had elevated adjusted odds of SNI as compared with infants born to white mothers under both definitions (Most inclusive: AOR_{FN} =3.05; 95%CI: 1.66, 5.62; Least inclusive: AOR_{FN} = 2.84; 95%CI: 1.08, 7.44). However, other or unknown ethnicity was associated with SNI under the most inclusive definition ($AOR_{Oth/Unk}$ = 1.45; 95%CI: 1.04, 2.04) but not the least inclusive definition ($AOR_{Oth/Unk}$ = 1.24; 95% CI: 0.69, 2.24).

Maternal employment status also showed differences in statistical significance across definitions; both levels of maternal employment status were significantly associated with SNI compared with employed mothers under the more inclusive SNI definition ($AOR_{student/homemaker}$ = 1.56, 95%CI: 1.14, 2.14; $AOR_{unemployed/other}$ = 2.17; 95%CI: 1.35, 3.49), while maternal employment status was not significantly associated with SNI as defined by the least inclusive SNI definition ($AOR_{student/homemaker}$ = 1.38, 95%CI: 0.80, 2.39; $AOR_{unemployed/other}$ = 1.06, 95%CI: 0.48, 2.34).

Antenatal steroid exposure and gestational age (as a continuous variable) were also significantly associated with SNI under the most inclusive definition ($AOR_{steroids}$ = 0.62; 95%CI: 0.39, 0.98 and AOR_{GA} = 0.82; 95%CI: 0.74, 0.92), but neither of these two factors was associated with SNI under the least inclusive definition of SNI ($AOR_{steroids}$ = 0.83; 95%CI: 0.40, 1.73, AOR_{GA} = 0.88; 95%CI: 0.73, 1.06).

Despite the above-mentioned differences in statistically significant associations between these determinants and SNI as defined by the most and least inclusive definitions of SNI, the AORs (expressing the strength of the association between each determinant and SNI under the 2 definitions) were not significantly different from each other. The confidence 95% confidence intervals of the AORs under the least inclusive definition included the point estimates of the most inclusive definition's AORs, indicating that the AORs were not significantly different at the 5% level

Maternal drug use during pregnancy, male sex, SNAP-II score >20, late onset sepsis, BPD, and brain injury (IVH ≥ grade 3 or PVL) consistently showed increased adjusted odds of SNI under both definitions of SNI, but the strength of the association varied. Adjusted odds ratios for these factors were generally higher when the least inclusive definition of SNI was used, although the 95% confidence intervals were wider for these estimates.

Goodness-of-fit was adequate for all models, as evidenced from the Hosmer-Lemeshow tests (p-HL). C-statistics (c-stat) showed an increase in classification accuracy of the model in regression steps 1-3 (increased value

towards 1) and a comparable accuracy between models in step 3 and the final models. The accuracy was better for the model using the least inclusive definition of SNI (0.82 vs 0.77 for the final models under the least and most inclusive definitions of SNI, respectively).

Table 8: Multivariable logistic regression model: Odds ratio and 95% CI for severe neurodevelopmental impairment (SNI) under the most inclusive definition of SNI

Variable	Level	Step 1	Step 2	Step 3	Final
Ethnicity	White	ref	ref	ref	ref
	Asian	1.08 (0.68, 1.70)	0.98 (0.61, 1.57)	1.11 (0.66, 1.86)	1.11 (0.66, 1.87)
	First nations	2.01 (1.13, 3.55)	2.28 (1.28, 4.08)	2.95 (1.60, 5.46)	3.05 (1.66, 5.62)
	Other/Unknown	1.30 (0.96, 1.76)	1.33 (0.97, 1.82)	1.50 (1.07, 2.10)	1.45 (1.04, 2.04)
Education	≥ High school	ref			
	< High school	1.03 (0.68, 1.56)			
Employment	Employed	ref	ref	ref	ref
	Student or homemaker	1.59 (1.19, 2.12)	1.61 (1.20, 2.06)	1.56 (1.14, 2.14)	1.56 (1.14, 2.13)
	Unemployed/other	2.05 (1.33, 3.16)	1.84 (1.18, 2.87)	2.08 (1.29, 3.35)	2.17 (1.35, 3.49)
Single parenthood	no	ref			
	yes	1.32 (0.80, 2.18)			
Maternal age	<25	1.12 (0.77, 1.63)			
	25-34	ref			
	35+	0.99 (0.71, 1.37)			
Parity	Primipara	0.82 (0.61, 1.09)			
	Multipara	ref			
Antenatal steroids	No	ref	ref	ref	ref
	Yes	0.52 (0.34, 0.79)	0.54 (0.35, 0.85)	0.64 (0.41, 1.03)	0.62 (0.39, 0.98)
Drugs	No	ref	ref	ref	ref
	Yes	2.56 (1.37, 4.80)	3.21 (1.72, 5.99)	3.31 (1.69, 6.48)	3.28 (1.68, 6.41)
Apgar <7	No		ref	ref	
	Yes		1.63 (1.24, 2.15)	1.30 (0.96, 1.75)	
GA	Continuous		0.71 (0.65, 0.79)	0.85 (0.76, 0.96)	0.82 (0.74, 0.92)
Sex	Female		ref	ref	ref
	Male		1.79 (1.36, 2.37)	1.89 (1.40, 2.55)	1.92 (1.42, 2.58)
Outborn	No		ref		
	Yes		0.97 (0.58, 1.63)		
SNAP-II>20	No			ref	ref
	Yes			1.53 (1.12, 2.10)	1.65 (1.21, 2.25)
Late onset sepsis	No			ref	ref
	Yes			1.52 (1.12, 2.07)	1.56 (1.15, 2.11)
BPD	No			ref	ref
	Yes			1.47 (1.08, 2.00)	1.54 (1.13, 2.09)
ROP grade≥3	No			ref	
	Yes			1.44 (0.97, 2.13)	
NEC	No			ref	
	Yes			1.20 (0.71, 2.02)	
IVH grade≥3	No			ref	ref
	Yes			4.72 (3.29, 6.77)	4.80 (3.35, 6.87)
		Step 1	Step 2	Step 3	Final
AIC		1525	1427	1280	1281
p-HL		0.53	0.20	0.36	0.06
c-stat		0.63	0.71	0.78	0.77

Table 9: Multivariable logistic regression model: Odds ratio and 95% CI for severe neurodevelopmental impairment (SNI) under the least inclusive SNI definition

Variable	Level	Step 1	Step 2	Step 3	Final
Ethnicity	White	ref			
	Asian	1.19 (0.51, 2.75)			
	First nations	1.94 (0.74, 5.09)			
	Other/Unknown	0.95 (0.51, 1.78)			
Education	≥ High school	ref			
	< High school	0.70 (0.31, 1.63)			
Employment	Employed	ref			
	Student or homemaker	1.63 (0.94, 2.81)			
	Unemployed/other	1.00 (0.37, 2.72)			
Single parenthood	No	ref			
	Yes	2.24 (0.99, 5.03)			
Maternal age	<25	0.77 (0.37, 1.61)			
	25-34	ref			
	35+	0.58 (0.29, 1.18)			
Parity	Primipara	0.65 (0.37, 1.16)			
	Multipara	ref			
Antenatal steroids	No	ref			
	Yes	0.53 (0.25, 1.11)			
Drugs	No	ref	ref	ref	ref
	Yes	3.33 (1.22, 9.12)	4.12 (1.64, 10.3)	3.77 (1.30, 10.9)	3.59 (1.25, 10.3)
5- minute Apgar <7	No		ref	ref	
	Yes		1.86 (1.11, 3.13)	1.44 (0.82, 2.53)	
GA	Continuous		0.68 (0.57, 0.81)	0.93 (0.75, 1.16)	
Sex	Female		ref	ref	ref
	Male		1.72 (1.02, 2.91)	1.79 (1.02, 3.13)	1.77 (1.01, 3.09)
Outborn	No		ref		
	Yes		1.93 (0.91, 4.10)		
SNAP-II>20	No			ref	ref
	Yes			2.10 (1.18, 3.73)	2.46 (1.42, 4.26)
Late onset sepsis	No			ref	ref
	Yes			2.00 (1.14, 3.52)	2.22 (1.29, 3.81)
BPD	No			ref	ref
	Yes			2.51 (1.35, 4.67)	2.86 (1.57, 5.21)
ROP grade≥3	No			ref	
	Yes			1.64 (0.86, 3.14)	
NEC	No			ref	
	Yes			1.24 (0.51, 3.01)	
IVH grade≥3	No			ref	ref
	Yes			6.22 (3.53, 10.9)	6.78 (3.90, 11.8)
		Step 1	Step 2	Step 3	Final
AIC		562	530	455	453
p-HL		0.78	0.86	0.77	0.45
c-stat		0.67	0.71	0.83	0.82

Table 10: Multivariable logistic regression model: Odds ratio and 95% CI comparing full set of variables associated with severe neurodevelopmental impairment (SNI) under either definition of SNI (most inclusive or least inclusive)

Variable	Level	Most inclusive Final extended	Least inclusive Final extended
Ethnicity	White	ref	ref
	Asian	1.11 (0.66, 1.87)	1.32 (0.53, 3.30)
	First nations	3.05 (1.66, 5.62)	2.84 (1.08, 7.44)
	Other/Unknown	1.45 (1.04, 2.04)	1.24 (0.69, 2.24)
Employment	Employed	ref	ref
	Student or homemaker	1.56 (1.14, 2.13)	1.38 (0.80, 2.39)
	Unemployed/other	2.17 (1.35, 3.49)	1.06 (0.48, 2.34)
Antenatal steroids	No	ref	ref
	Yes	0.62 (0.39, 0.98)	0.83 (0.40, 1.73)
Drugs	No	ref	ref
	Yes	3.28 (1.68, 6.41)	3.56 (1.42, 8.92)
GA	Cont.	0.82 (0.74, 0.92)	0.88 (0.73, 1.06)
Sex	Female	ref	ref
	Male	1.92 (1.42, 2.58)	1.72 (1.03, 2.89)
SNAP-II>20	No	ref	ref
	Yes	1.65 (1.21, 2.25)	2.04 (1.21, 3.44)
Late onset sepsis	No	ref	ref
	Yes	1.56 (1.15, 2.11)	1.79 (1.07, 2.98)
BPD	No	ref	ref
	Yes	1.54 (1.13, 2.09)	2.38 (1.36, 4.17)
IVH grade≥3	No	ref	ref
	Yes	4.80 (3.35, 6.87)	5.54 (3.27, 9.39)
		Final extended	Final extended
AIC		1281	545
p-HL		0.06	0.12
c-stat		0.77	0.81

NOTE: Abbreviations in Tables 8,9 & 10: GA (gestational age), SNAP-II (Score for Neontatal Acute Physiology - 2nd edition), BPD (bronchopulmonary dysplasia), ROP (retinopathy of prematurity), NEC (necrotizing enterocolitis), IVH (intraventricular hemorrhage); AIC (Akaike information criterion); p-HL denotes p-value for Hosmer-Lemeshow tests; c-stat denotes C-statistics (area-under-the-curve).; Infants with missing ROP were considered ROP negative.

4.3.4 Additional analyses

4.3.4.1 Site differences

Analysis of individual CNFUN sites showed statistically significant differences in incidence rates of SNI between sites depending on the definition of SNI used to classify infants with severe neurodevelopmental impairment (Table 11). Incidence of SNI was significantly elevated under the most inclusive SNI definition compared with the least inclusive definition for NICU sites with at least 50 infants with complete follow up (n= 14). Incidence rates of SNI ranged from 4.4% to 32.7% under the most inclusive definition and 0% to 7.7% under the least inclusive definition. Comparing rates of SNI under the two definitions showed that the site-specific risk ratio for SNI (under the most inclusive vs. least inclusive definition) ranged from 2.79 (95%CI: 1.56, 4.97) to 10.0 (95% CI: 1.31, 76.1).

Ranks of these CNFUN sites by SNI incidence rates were inconsistent, depending on which definition of SNI was used. Only sites L and J maintained their rank across definitions (9^{th} and 14^{th}, respectively). The other 12 sites (86%) changed rank by at least one position. The greatest change was observed for sites U and X; both moved 5 positions in the rankings (13^{th} or 8^{th} and 7^{th} or 12^{th}, respectively). Rankings and the magnitude of change in rank between the most inclusive definition and the least inclusive definition is shown in Table 12.

Table 11: CNFUN site comparison of severe neurodevelopmental impairment (SNI) incidence for sites with ≥50 subjects (n=14)

Site	Most inclusive SNI definition (%)	Least Inclusive SNI definition (%)	Relative Risk	95% Confidence Interval
A	7/158 (4.4)	2/158 (1.3)	3.50	0.74, 16.59
B	8/109 (7.3)	0/109 (0)	∞	–
E	22/210 (10.5)	6/210 (2.9)	3.67	1.52, 8.86
F	39/199 (19.6)	14/199 (7.0)	2.79	1.56, 4.97
H	25/124 (20.2)	5/124 (4)	5.00	1.98, 12.64
J	17/52 (32.7)	4/52 (7.7)	4.25	1.53, 11.78
K	21/171 (12.3)	4/171 (2.3)	5.25	1.84, 14.97
L	13/80 (16.3)	3/80 (3.8)	4.33	1.28, 14.63
N	8/92 (8.7)	1/92 (1.1)	8.00	1.02, 62.69
Q	44/242 (18.2)	10/242 (4.1)	4.40	2.27, 8.54
R	10/71 (14.1)	1/71 (1.4)	10.0	1.31, 76.08
U	13/64 (20.3)	2/64 (3.1)	6.50	1.53, 27.65
X	18/122 (14.8)	6/122 (4.9)	3.00	1.23, 7.30
Z	33/221 (14.9)	5/221 (2.3)	6.60	2.63, 16.59

Table 12: CNFUN sites ranked by decreasing incidence of severe neurodevelopmental impairment (SNI) under the most and least inclusive definitions of SNI for sites with ≥50 subjects (n=14)

Site	Rank Using Most Inclusive SNI	Rank Using Least Inclusive SNI	Rank Change
A	1	3	-2
B	2	1	+1
N	3	2	+1
E	4	7	-3
K	5	6	-1
R	6	4	+2
X	7	12	-5
Z	8	5	+3
L	9	9	0
Q	10	11	+1
F	11	13	-2
H	12	10	+2
U	13	8	+5
J	14	14	0

4.3.4.2 Missing values

Though 2187 infants were included in the study, outcome data were not available for some SNI categories for a number of infants. Table 13 details the number of infants with missing data in each domain, as well as the number with missing data in a specific domain plus one or more other domains. For example, 265 infants had missing information for the Bayley-III motor score, and 131 of these also had missing information in 2 other domains. A large portion of this group was probably not assessed with the Bayley-III at all, which is evidenced by the fact that 151 infants in the study were missing a Bayley-III cognitive score, and 234 were missing a Bayley-III language score. This suggests that if an infant was not assessed using the Bayley-III, they would have missing data for all three SNI domains.

Table 13: Subjects missing data in one or more domain using CNFUN definition of severe neurodevelopmental impairment (SNI) (≥1 of: GMFCS 3-5; Bayley-III <70 for motor, cognitive, or language composite score; hearing aids or cochlear implants; bilateral blindness)

	Number of Subjects Missing Multiple Domains (n)					
SNI domain	Total Missing	No other missing	1 other missing	2 other missing	3 other missing	4+ other missing
CP (GMFCS)	44	9	6	18	9	2
Motor (Bayley)	265	70	40	131	22	2
Cognitive (Bayley)	151	6	2	121	20	2
Language (Bayley)	234	45	39	127	21	2
Hearing	73	22	21	19	10	1
Vision	174	116	28	19	10	1

Table 14: Subjects missing data by domain using CNFUN definition of severe neurodevelopmental impairment (SNI) (≥1 of: GMFCS 3-5; Bayley-III <70 for motor, cognitive, or language composite score; hearing aids or cochlear implants; bilateral blindness)

		Number of Subjects Missing Multiple Domains (n)					
SNI domain	Total (n)	No missing	1 missing	2 missing	3 missing	4 missing	5 missing
CP (GMFCS)	48	12	11	1	20	4	0
Motor (Bayley)	93	67	24	1	1	0	0
Cognitive (Bayley)	59	29	25	4	0	1	0
Language (Bayley)	188	130	49	4	4	1	0
Hearing	56	31	9	3	12	1	0
Vision	32	12	9	1	10	0	0

The number of subjects with missing information among infants with overlapping impairments can be seen in Table 14. For instance, of the 48 infants who had severe cerebral palsy, 12 had complete data. The majority of subjects included in this study had complete information on all SNI domains (n=1681; 58.6%); however, among subjects who had one or more severe impairments, a smaller proportion had complete data. Complete data was available for 25% of infants with CP, 72% with motor impairment, 49% with cognitive impairment, 69% with language impairment, 55% requiring cochlear implants or hearing aids, and 38% of bilaterally blind infants.

5. Discussion

5.1 Comments on study results

The aim of this study was to identify existing definitions of severe neuro-developmental impairment, and to examine the effect of varying definitions on the incidence rate of SNI and the strength of association between known risk factors and SNI in the very preterm infant population.

Seven distinct definitions of composite severe neurodevelopmental impairment used in the literature that met our inclusion criteria were identified. These definitions included severe impairment in most or all of the following six domains but with different thresholds: cerebral palsy, Bayley-III motor, Bayley-III cognitive, Bayley-III language, vision and hearing.

Results show a marked difference between the SNI incidence rate calculated using the definition with the broadest inclusion criteria (Def. 1) and the definition with the most restrictive criteria (Def. 7). The former yielded a SNI rate 329% higher than the latter (14.9% vs. 3.5%). Such a remarkable difference in incidence rate raises concerns about interpretation of results in the literature, especially if the readers fail to appreciate difference in prevalent SNI definitions. This issue is significant because decision making about critically important matters such as the withdrawal of life-saving care in the NICU takes place based on the probability of severe neurodevelopmental impairment. A 3.5% probability of SNI is substantially different from a 15% probability of SNI, and this difference can significantly affect women in preterm labour at very early gestation who may have to decide between active resuscitation and palliative care for their very

preterm newborn infant. Similarly, eligibility for early intervention, rehabilitation or other health care support for the VPT survivor may be affected by the definition/criteria for SNI used.

Risk factors that were statistically significantly associated with SNI varied depending upon SNI definition. Maternal ethnicity, employment status, antenatal steroids treatment, and infant gestational age (GA) showed inconsistency in the statistical significance of the association with SNI, depending on the definition.

While the odds ratios were reasonably similar when using varying definitions, “statistical significance” of some risk factors varied. For example, the adjusted odds of SNI for gestational age was 0.82 under the more inclusive definition and 0.88 under the less inclusive definition. The reason for a disparity in statistical significance despite similar AORs differences is the precision of the two estimates. The less inclusive definition had fewer infants classified as SNI and therefore the 95% confidence intervals around the point estimate were wider and the power to detect a significant association was lower.

There are several potential reasons why some risk factors show statistically significant associations with the outcome under one definition but not the other. First, the higher rate of the outcome (given a fixed study size) when using a more inclusive SNI definition results in an increased statistical power (i.e., and a greater ability to rule out the null hypothesis if a true association exists between the risk factor and the outcome). This higher rate of the outcome is also responsible for greater precision of the effect estimate (i.e., the confidence intervals for the odds ratio are narrower). The need for greater power is

especially evident for factors that are less strongly associated with SNI, which could explain why social determinants of health such as maternal ethnicity and employment status appeared to be significantly associated with SNI under the broader definition. Second, it is possible that factors that are directly related to SNI may not display statistical significance in a logistic regression model due to collinearity. For instance, gestational age and birth weight are highly correlated, and could potentially disrupt regression modeling if both are included in the model. It is possible that low gestational age affects a broad range of neurodevelopmental impairments, and therefore displays a significant association with the more inclusive SNI definition. Third, some prenatal and neonatal conditions may be associated predominantly with less severe impairment; these conditions are more likely to be recognized when using the more inclusive SNI definition. Fourth, this study population includes only surviving infants, therefore risk factors that are a common cause of death in VPT infants may show a weaker association with SNI (survival bias). In addition, factors that influence access to health care services and specialist visits may influence the association with SNI. For example, single parents or non-white mothers (First Nation or other ethnicities) may have barriers to access specialized health care interventions for their infants in the first two years after birth (87–90), which may result in a higher rate of SNI (91), especially when the more inclusive SNI definition is used. In summary, variables that displayed statistical significance under both definitions were those that were more strongly associated with adverse neurodevelopmental outcomes in general. Variables

less strongly associated with SNI showed variable significance in their associations with SNI because the precision of the estimates of effect was altered by the varying incidence of SNI under the different definitions. Regardless of the reasons for the disparity in statistical significance for risk factors across SNI definitions, this issue should raise a concern about how we interpret scientific literature and report results.

Site-specific rates of SNI were significantly higher when using the more inclusive definition for all but one of the fourteen sites (with ≥50 subjects). This suggests that even at the individual site level, different definitions of SNI can impact reported results, and could thus have an impact on decision making in the NICU. Incidence rates at individual sites ranged from 4.4% to 32.7% under the more inclusive definition. A 7.5-fold difference in rates suggests some sites may be more or less successful in preventing severe long-term neurodevelopmental impairment than others. Similarly, under the less inclusive SNI definition, incidence rates ranged from 0% to 7.7%. However, for meaningful site comparisons, these rates would need to be adjusted for baseline characteristics of admitted infants in order to make appropriate comparisons. Ranking sites by SNI incidence rates also showed inconsistencies dependent upon the definition of SNI used. Most of the sites (86%) changed positions by at least one rank, and two sites changed their standing dramatically, moving 5 positions. This can have implications for inter-network benchmarking, because determining which sites lead and which sites lag behind in terms of low SNI incidence becomes dependent on SNI definition.

Comparisons between SNI rates reported in the literature and CNFUN-based results are also important. While most incidence rates reported in the literature were comparable with our CNFUN rates, there were some exceptions. Definitions that displayed similar rates between original and CNFUN data were from the Australian and New Zealand Neonatal Network (ANZNN) from the years 2013 and 2012, respectively (in reference to Table 6).(30,45) CNFUN calculated rates and ANZNN 2013 calculated rates of SNI were 4.1% and 4.6%, respectively; these rates were both 3.5% under the definition used by ANZNN in 2012.(31,46) In contrast, the Victorian Infant Collaborative Study's definition (VICS) in the CNFUN dataset had an SNI incidence rate of 8.4% compared to the original 3.7% incidence rate that Doyle et al. reported in their study.(59) This regional Australian network has an incidence rate that is closer to the incidence in the larger ANZNN network. These findings can be partially explained by Australian Bayley-III scores that are generally higher than Bayley-III scores reported in Canada or the US, as Australian networks use a population control comparison group rather than the American normed standardized cohort that was initially used to develop the Bayley-III.(59,92)

The United States' NICHD network reported rates of severe neurodevelopmental impairment of 13.7% and the Swedish EXPRESS group reported rates of 11.0%, which are substantially higher than the incidence rates calculated from CNFUN data using the same definitions (7.8% and 3.9%, respectively).(42,52) The NICHD and EXPRESS studies included infants <27 weeks gestation, which encompassed more vulnerable infants as compared with

our CNFUN study (infants born at 23-28 weeks). This may explain the discrepancies, since increasing gestational age coincides with increased odds of survival and decreased odds of severe neurodevelopmental impairment.(42)

Direct comparisons of incidence rates of SNI between follow-up networks using the same SNI definition would facilitate the identification of populations, intervention strategies, and other factors that may be associated with lower rates of adverse neurodevelopmental outcomes. Networks could pool the data and examine successful strategies from each network to enhance quality improvement and collaborative research to expand our understanding of neurodevelopmental impairment.

5.2 Rationale for different definitions of severe neurodevelopmental impairment

If varying definitions among countries or networks can lead to drastically different incidence rates that prevent direct international comparisons, why do different definitions exist at all?

The decision about specific definitional criteria for developmental outcomes is often a value judgment on the part of the investigators with no reason given. Some investigators may value the positive outcomes of prematurity and choose a restrictive definition that identifies only the most severely impaired and hence low SNI rates. However, for children and their families' seeking community services, a less restrictive definition may be in their best interest. This option may help raise awareness for the less severe and less visible impairments.

The British Association of Perinatal Medicine, articulated their intent when they defined their definition of SNI: “useful as a descriptor of outcomes which were likely to impair independence throughout childhood and might be useful to parents in the process of perinatal decision making” for instance, aimed at a restrictive definition to maximize utility for counseling parents of very preterm infants through the early neonatal period when they may have to make critical decisions regarding the continuation or withdrawal of life support.(58) Other networks may wish to either identify criteria for clinical screening with a high specificity or to meet more restrictive diagnostic criteria with a high sensitivity. Furthermore, the definition of SNI may be crafted for the sake of calculating and comparing prevalence rates across time and place. A follow-up network may decide to use a specific definition because it is consistent with how it has been defined in the past – allowing for a comparison of temporal trends within the network. This is valuable from an epidemiologic point of view because the temporal trend in the incidence and prevalence of SNI in a specific population can inform intervention programs and identify avenues for improvement.

Another factor that could contribute to definition variance is the fact that developmental tests are updated periodically to address various shortcomings or gaps in assessment and to remain valid and relevant. Various modifications may be made to avoid, for example, the Flynn effect, a phenomenon whereby the collective performance on any IQ test improves over time across successive birth cohorts.(93) As a result, developmental tests have to be periodically recalibrated to adjust the average score. This can cause problems for researchers studying

longitudinal cohorts where different editions of a developmental or intelligence test were used. For instance, in 2006 the Bayley Scales of Infant and Toddler Development-2nd Edition (BSID-II) was updated and released as the Bayley-III. However, it has been shown that the Bayley-III underestimates the level of impairment compared to BSID-II.(92) For this reason, some studies have altered their criteria for severe impairment to correct for this underestimation. Moore *et al.* showed that by moving the cutoff to a Bayley-III score <80 instead of <70, the sensitivity of the test markedly improved without sacrificing specificity.(94) The Neonatal Oxygenation Prospective Meta-analysis (NeOProM), Canadian Oxygen Trial (COT), and Benefits of Oxygen Saturation Targeting (BOOST-II) trials were parallel RCT studies that aimed to determine the appropriate oxygen saturation level for very preterm infants (<28 weeks gestation) in the NICU.(54,95) These studies used a BSID-II cutoff of <-2SD below the mean to calculate their sample size. However, once data was published showing that the Bayley-III underestimates the degree of impairment, the criteria were adjusted to include all scores below <1SD to define severe impairment (Bayley-III scores <85) to maintain the proportion of the cohort anticipated to have a severe impairment. As these concurrent oxygenation studies were randomized control trials and not neonatal follow-up network reports, this SNI definition (<1SD) was not included in our study – if it were, the incidence rate of SNI would have increased to 46% (55), making the highest incidence rate more than 13 times higher than the lowest rate in our study.

The evolution of developmental assessment tools often leads to discussions about how severe impairment should be defined every time an updated edition of the assessment test is released. Such evolution can lead to a large array of different definitions; some groups may chose to adjust for under- or overestimation of severe impairment each time a new version of the test is released, while other groups may elect to use a consistent cut-off value across all editions (e.g., <2SD). The latter option acknowledges that whereas the severity of impairment may change with each new/recalibrated test, studies will keep proportions of impaired infants constant (by consistently reporting those with the lowest 2.5% on the test scores). These decisions stem from the value judgments about the intent or purpose of a network's definition.

Choosing which criteria to include in the definition of severe impairment may also simply depend upon what is realistically feasible for a particular network. A standardized neurodevelopmental exam requires a fully trained examiner available at a designated location, and the test can take more than one hour to be properly administered. This may create budgetary, time, geographic, and other constraints. For instance, a network may choose to use the GMFCS cerebral palsy severity score as a measure of motor function in lieu of the Bayley-III motor composite score if they can only afford to perform the language and cognition portions of the Bayley-III. This may explain why some networks in this study included only 2 of the 3 composite scores, or excluded a measure of hearing impairment from the SNI definition. Another consideration is the financial and logistic ramifications of a broader definition of SNI that may require

increased funding for special care programs and support for VPT survivors. A country or region with scarce resources may opt for a narrower definition so that only the most severely impaired infants who require the most intervention and therapy throughout their life-course are targeted.

Finally, differing definitions of SNI may reflect the debate about what should be considered the normal range of neurodevelopment, i.e., what should be considered severe enough impairment to be regarded as 'outside of the normal developmental range'. When using developmental tests such as the Bayley-III, some researchers have opted for criteria that consider any score >2SD below the mean as a severe impairment. Assuming the test norms reflect the population being evaluated, this identifies 2.5% children with the lowest scores as severely impaired, irrespective of the impact of the impairment on the child. Other groups argue that this cutoff is not restrictive enough to limit the categorization of SNI to those children with the greatest impact of the impairment on the child, and have chosen a cutoff >3SD below the mean. This very restrictive cutoff, one might argue, will capture severe impairment more accurately, though it may exclude some children with severe impairment. This ongoing debate about what should constitute the typical range of neurodevelopment will continue to shape the criteria for the definition of SNI and thus impact the incidence reported in annual reports or the results of studies specific to each network. Attempts to create a rigid dichotomy between normal vs. impaired will pose a challenge because neurodevelopment outcomes are heterogeneous and include a wide range of functional outcomes.

5.3 Sources of variation in SNI outcome between studies

It is important to note that there may be several sources of variation in the data used to define SNI. Variation in SNI data may explain some of the variation in incidence of severe neurodevelopmental impairment across sites and networks. Some possible sources of variation include: 1) population differences, 2) variation in inter-rater reliability, 3) temporal changes in mean standardized scores, 4) child's age at assessment, 5) limitations of assessment methods, and 6) variation in utilization, type and quality of post-NICU intervention programs .

1) Differences in sociodemographic and behavioral characteristics, or geographical barriers to access to care may contribute to differences in risk of adverse outcomes, and thus contribute to differences in incidence of SNI. These differences are typically addressed in multivariable analyses in an attempt to mitigate the effect of population differences, however it is possible that residual confounding biases the results when different populations are compared.

2) Administering developmental tests requires trained professionals who have access to the resources and knowledge necessary to perform the examination. While standardized training in administering and interpreting assessment tools such as the Bayley-III is required, there is potential for interrater differences in scoring and interpretation. Similarly, the assessment is performed at one point in time which may not reflect the child's usual development (e.g.,feeling ill, tired, new environment, etc.).

3) The Flynn Effect may also contribute to variance in the data used to define SNI because children's cognitive scores have been shown to improve over time.

As a result standardized developmental assessment tools are revised regularly. Temporal comparisons are therefore complicated by both the Flynn effect and new versions of assessment tools. The validity of comparing incidence of SNI between two studies that differ by several years may be reduced as a result of this phenomenon.

4) As children age they are continuously maturing and developing new skills. Developmental tests calculate ability by age windows. Systematic differences in whether children are evaluated in the early or late part of the window could create a systematic bias.

5) Furthermore, whenever an assessment tool is used in children who differ from the population in which the tool was validated the results can vary. For example, assessments such as the Bayley-III were validated in English speaking children but is sometimes administered to non English speaking children. Similarly, if a child has a hearing or visual impairment, inability to perform a task may be interpreted incorrectly.

6) Early intervention programs may affect development status and consequently the classification of severity of neurodevelopmental impairment. This study did not include interventions between NICU discharge and 18-21 month follow-up. These interventions are more proximal to the outcome, and could have a significant impact on the developmental trajectory of the VPT survivor. Previous studies suggest that a variety of early intervention programs with parent involvement are successful in improving long-term

neurodevelopmental outcomes in this population, with a peak effect at 36 months corrected age.(96)

The predictive validity of developmental assessments at 18-24 months is limited in terms of future intelligence and performance. Children are tested at 18-24 months corrected gestational age because this is considered the earliest time point at which a major disability can be reliably diagnosed. Reports of the predictive validity of the 2nd and 3rd editions of the Bayley were inconsistent for classifying impairment later in life. Some studies have shown strong correlations with early- and later-life impairment (97), with 81% of infants retaining their classification at age 4, while other studies reported poor predictive validity, with a sensitivity less than 38% for cognitive impairment and less than 47% for language impairment at four years of age.(33,98) A meta-analysis showed that Bayley-III motor scores explain only 12% of variance in motor function later in life.(99) In other words, when using Bayley-III for neurodevelopmental assessments at 18-24 months of age, 88% of the variance in impaired motor function later in life remains unexplained. As a result, predicting the impact of an SNI classification on an infant's long-term neurodevelopmental functioning remains a significant challenge, and different networks may choose different cutoff criteria for SNI with the intention of targeting a definition with higher sensitivity or predictive validity for severe impairment later in life.

5.4 Limitations of SNI measurement and interpretation

Binary categorization (e.g., impaired vs non-impaired) is inherently difficult in the sense that impairment and normal development exist on a spectrum. Composite binary outcomes are often derived from components that are continuous measures. The reclassification, for example from a scalar Bayley-III score into a binary outcome, causes a loss of information. Those with a Bayley-III score of 69 are not much different from those with a score of 71, but with a cutoff of 70, one will be described as impaired and the other as normal. Composite outcomes have been described as having utility in increasing statistical precision, paired with a consequential decrease in the uncertainty surrounding the effect estimate (100), as the individual components of the composite outcomes may not have an equal impact on health (or function). Similarly, in our study, severe hearing impairment may not be functionally equivalent to severe cerebral palsy, which may not be equivalent to a Bayley-III score of 69, etc., as the impact of each impairment domain can vary. Despite this limitation, composite outcomes are often used because they provide a single outcome variable of interest. A composite outcome is advantageous in research studies when it is necessary to identify a primary outcome and it is not evident what single clinical measure is the most effective in capturing the essence of impairment. It is often impractical to investigate a multitude of factors that together make up an SNI composite outcome, and the overall message and purpose can get lost in cumbersome tables or complicated figures. Thus, composite variables are used with the understanding that some information that is confined in the complexity of multiple

measurements may be lost. Selection of the individual components of the composite outcome often depends on a consensus of researchers, clinicians, and other stakeholders. In this study, for example, we found that while some networks considered hearing loss to be a component of SNI, others did not, perhaps because they felt that hearing loss on its own is not severe enough to constitute a severe impairment.

The interpretation of composite outcomes can be challenging due to the aforementioned loss of information inherent in combining several outcomes. The SNI composite outcome captures impairment as a measure of body structure or function. This is in contrast with International Classification of Functioning, Disability, and Health that aims to identify limitations of participation in various normal life activities. For example, children treated with a cochlear implant may have a severe hearing impairment and meet the criteria for SNI, while they function with minimal limitation in day-to-day activities. This opens up a debate about meaningful health outcomes from parents' and providers' point of view. Parents and providers may perceive a child's health according to the level of restrictions to activities of daily living as more meaningful than the level of impairment. An infant may have a score <70 on the Bayley-III, but interpreting what this score *means* for parents and the child may not be straightforward. Counseling should come from trained professionals who are intimately familiar with how the developmental tests are administered, their predictive validity, and who are aware of limitations of a one-time assessment (of a potentially irritable or uninterested toddler, whose underperformance may lead to misclassification as a

neurodevelopmentally impaired child). This difference in the interpretation of impairments vs functional ability can have important consequences for parents and children.

Furthermore, any meaning drawn from a classification of severe neurodevelopmental impairment must be viewed through a contextual lens of societal norms and expectations. If a VPT survivor has cerebral palsy and requires a wheelchair for transportation in an industrialized country, that person can still participate in most day-to-day activities. This may not be the case for a child with cerebral palsy in a low-income country. The cultural or social context of impairment, albeit difficult to capture due to its complexity, is often overlooked when creating clinically oriented outcome definitions.

5.5 A standard definition

Though there are many considerations for choosing specific criteria for a SNI definition, our results show that a multiplicity of definitions can lead to miscommunication and confusion in clinical practice and research. Moving toward a standardized, universal definition of neurodevelopmental impairment could help address this problem.

The first reason for agreeing upon an international standardized definition for severe neurodevelopmental impairment is for a clarification of terms. A universal definition would allow direct spatial and regional comparisons. In addition to reporting results using the standard definition, each regional network can continue using their current definition, if different. This course of action may

also encourage networks to justify an alternative definition that differs from the universal standard.

Some of the definitions that are currently used to classify infants as severely neurodevelopmentally impaired are based on structural factors (as determined by medical diagnostic tests such as radiologic imaging) rather than function, which may not be optimal. Health care providers should be involved in the creation of a standardized definition to ensure the definition is capturing information that is valuable from a clinical perspective, but also from a counseling and planning perspective. A proper definition should provide parents with an understanding of what it *means* for their child to grow up living with impairment. End users, such as patients and their families, may find measures of impact on activities of daily living, restrictions on their ability to participate socially and overall quality of life more meaningful. It may be the case that a standardized definition supplements or even replaces some of the common measures of SNI with other measures, such as a prediction of social dysfunction, level of independence, or necessity of life-long care. However, valid measures of these outcomes are very limited for children less than 3 years of age.

Using a multilevel categorical system rather than dichotomous approach to define impairment may also improve interpretation. Some studies already use more than one category of impairment, such as “severe impairment” and “moderate impairment”. A tiered system could also be created that attempts to rate an infant’s impairment on a scale (e.g., from 1-10). This scale may factor in the number of domains affected and the severity of impairment in each domain.

5.6 Strengths and Limitations

5.6.1 Strengths

This study provides a detailed review of current SNI definitions used by neonatal networks in a number of industrialized countries. The strength of this study includes a large, nationally representative cohort of very preterm infants that was used to demonstrate the impact of different severe neurodevelopmental impairment definitions on rates of SNI and associations between common risk factors and SNI. CNFUN is a relatively large network and several measures have been employed to assure good quality of collected data (e.g. cross-referencing CNN and CNFUN data using unique patient identifiers during data linkage, and vetting information after obtaining data from each participating site).(57)

5.6.2 Limitations

There were some limitations in this study. Firstly, some measures in CNFUN may not have enough detail to be directly compared with similar measures from other networks. For example, some networks defined visual impairment using a specific measure of visual acuity (e.g., <6/60), whereas CNFUN used simple categories such as ‘normal visual function’, ‘blind in one eye’, or ‘bilaterally blind’, with ‘bilateral blindness’ as the threshold for severe impairment. Potential non-differential misclassification rising from such

categorization could have potentially biased our results towards the null; if CNFUN had a more detailed assessment of vision to match other networks, the differences in SNI definitions and incidence rates may have been even greater.

Another limitation was the method of ascertainment of vision and hearing outcomes, which were performed either via parental questionnaire or by using results from hearing and vision tests before the 18-21 month follow-up visit at CNFUN. The former may have introduced bias by relying upon parental knowledge of medical diagnoses and the willingness to share such information with CNFUN. Consequently, it is possible that some infants were misclassified in their visual or hearing impairment.

As with most studies, missing data was present and may have adversely affected results. Among several children who were impaired in one or more of the six CNFUN criteria for SNI, data regarding impairment in other domains were missing. For instance, of 48 infants with cerebral palsy, 26 (54%) had missing data for the Bayley-III Language component. It is likely that these infants were too severely impaired due to their severe cerebral palsy to complete the Bayley-III test. However, the specific reasons for missing data are not known. Information on some risk factors was missing in a substantial proportion of children data; for example, information on clinical chorioamnionitis was missing in 15.9% of cases and information on retinopathy of prematurity was missing in 20% of children. Information on some risk factors was incomplete or not available, e.g., maternal body mass index. However, this was a pragmatic study using data that were collected with the best intentions for complete follow-up and data

collection given a finite budget. Other national and regional neonatal follow-up networks likely experience similar issues with missing data on neurodevelopmental assessment. Also, it was not our goal to provide the most accurate rate of SNI; our objective was to demonstrate the differences in SNI rates when various different SNI definitions are used.

5.7 Recommendations

This study underscores the need for education about the effect of outcome definition on the results, including the impact on incidence rates and the strength of association between risk factors and the outcome. Readers should be cognizant of this effect when interpreting and making conclusions from the literature. Likewise, researchers need to explicitly define primary and secondary outcomes and need to consider the definitions used in previously published studies, especially if planning to compare results with those studies.

Furthermore, clinicians and researchers often use factors that are "statistically significant" in their association with the outcome to decide about preventive measures or to identify vulnerable groups. This study lends support to the many statisticians calling for an end to the use of p-values alone for decision-making.(101) We suggest a more judicious process involved in the interpretation of results. Rather than targeting factors that display statistical significance, a more nuanced method can be used which considers etiological factors that are biologically plausible, and factors that are clinically important. A statistical association can be described using various measures which provide more

information about the nature of the association (e.g., presenting Odds Ratio and 95% CI, 90% CI and 99% CI; rather than just a p-value <0.05).(101) Such reporting not only provides an indication of multiple alpha-error levels ($p<0.1$, $p<0.5$, $p<0.01$) but also a measure of the strength of the association and the precision of the point estimate. There is a role for education about the effect that imprecise statistics can have on interpretation of a study's results.

Moreover, the substantial differences in definitions of SNI should encourage discussion about creating a globally accepted uniform definition of severe neurodevelopmental impairment. This definition could be used in tandem with a network's previous definition to present rates of SNI, if the two definitions are not the same. A universal definition of SNI would provide clarity in terminology, and facilitate inter-network and global benchmarking. This report will not speculate about exactly which cut-off criteria should be the universal standard, but some characteristics of such a definition would include:

1. Multiple levels of SNI – to better represent the continuum of impairment, the definition should have at least 3 levels of severity (e.g., Moderate, Severe, Critical)
2. Representative of impairment severity – infants requiring cochlear implants often function well, especially compared with infants with multiple impairments; as such, the definition of SNI should accurately represent the impact on health and wellbeing.

3. Multiple viewpoints – physicians, parents, and the VPT survivors themselves may have a different understanding of “severe impairment,” and each perspective should be considered
4. An ability to reasonably predict future health status – some measure of predictability for life-long impairment would be valuable when classifying infants as severely impaired at 18-24 months
5. Flexibility – new versions of developmental tests are released periodically, and the definition of SNI should allow these updates to easily identify SNI.

6. Conclusions

The results of this study highlight the importance of careful examination of the outcome definitions used in studies, especially when composite outcomes such as SNI are reported. Readers should be aware that the definitions used in one published report may not be consistent with other studies examining the same outcome. Differing definitions of SNI lead to artefactual variations in the incidence rates of SNI and also affect the association between common risk factors and SNI. Although using the P-value < 0.05 as a criterion of statistical significance is attractive for its simplicity, it is preferable to rely on the strength of the association between a risk factor and the outcome, the precision of the estimate, biologic plausibility and the clinical relevance of the findings. Interpretation of SNI incidence rates and the association between risk factors and SNI should be carried out with a clear understanding of the definition of SNI. These findings provide support for the creation of an internationally standardized definition of severe neurodevelopmental impairment.

7. Future Directions

This study raises concerns about the inconsistent criteria used to define severe neurodevelopmental impairment and the results that stem from such variance. This report discussed potential reasons for the use of different SNI definitions, and made recommendations based on the reported findings. The next step forward may be a qualitative exploration of what severe impairment means to key stakeholders. Parents and VPT survivors may value the impact of an impairment on daily life differently than clinicians. Key stakeholders convening to start a dialogue about quality of life and the value judgments surrounding neurodevelopmental impairment will be a critical next step in taking a reasoned approach to the classification of impairment in very preterm infants. It is hoped that the product arising from such focus groups will be the identification of a standardized definition of severe neurodevelopmental impairment that is widely accepted by the pediatric follow-up community.

Similarly, there is a need to educate pediatricians and other caregivers about the interpretation of SNI definitions and their impact on the functioning and quality of life of children as they mature and grow and how to use this data to counsel parents about long-term prognosis.

Another avenue of research is a more detailed analysis of site-specific factors related to differences in SNI incidence rates. In this study there were significant differences in crude SNI rates across CNFUN sites. It would be informative to perform further logistic regression analyses to investigate variables associated with higher rates of SNI. Factors of interest may include health

service delivery factors such as physician:patient ratio and type of NICU (private room vs. bay), environmental factors such as NICU noise level, and individual level factors such as the amount of exposure to stressful procedures and dosing of opioid analgesia/anesthesia, as well as the other known risk factors included in this study. There is already evidence that site-specific factors can differentially affect the incidence of intraventricular hemorrhage, which is a strong predictor of SNI. NICU characteristics were responsible for 31% of the variation in IVH incidence, independent of other risk factors.(102) This suggests that there may be modifiable NICU factors that can reduce rates of adverse health outcomes and severe neurodevelopmental impairment among newborns. Such a prospect creates a strong impetus for further research examining site-level characteristics associated with SNI, and novel avenues to improve health and neurodevelopment among vulnerable infants.

This study identified several risk factors strongly associated with SNI. Quality improvement initiatives can be implemented to improve outcomes by focusing on modifiable risk factors. Antenatal steroid treatment was associated with decreased odds of SNI, meaning obstetrical practices could be improved upon by ensuring women in very preterm labour are treated with antenatal steroids. Modifications in neonatal practice may also reduce rates of SNI through prevention of critical illness in the NICU. Based upon the results of this study, optimal targets for improvement include reduction of sepsis, bronchopulmonary dysplasia, and intraventricular hemorrhage, which were all strongly associated

with SNI. Similarly, targeting societal risk factors such as unemployment and illicit drug use may be effective in reducing SNI rates in the VPT population.

References

1. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet. 2008 Jan;371:75–84.

2. Public Health Agency of Canada. Canadian Perinatal Health Report - 2008 edition [Internet]. 2008 [cited 2015 May 8]. Available from: http://www.phac-aspc.gc.ca/publicat/2008/cphr-rspc/index-eng.php

3. Statistics Canada. CANSIM - 102-4513 - Live births, weeks of gestation indicators, by characteristics of the mother and child, Canada [Internet]. 2015 [cited 2015 May 8]. Available from: http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=1024513&paSer=&pattern=&stByVal=1&p1=1&p2=-1&tabMode=dataTable&csid=

4. Clements KM, Barfield WD, Ayadi MF, Wilber N. Preterm birth-associated cost of early intervention services: an analysis by gestational age. Pediatrics. 2007;119(4):e866–74.

5. 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.

6. Trasande L, Malecha P, Attina TM. Particulate Matter Exposure and Preterm Birth: Estimates of U.S. Attributable Burden and Economic Costs. Environ Health Perspect. 2016 Mar;(ePub ahead of print).

7. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, et al. Born Too Soon: The global epidemiology of 15 million preterm births. Reprod Health. 2013 Nov 15;10(Suppl 1):S2.

8. Preterm Birth: Causes, Consequences, and Prevention [Internet]. Washington, D.C.: National Academies Press; 2007 [cited 2016 May 30]. Available from: http://www.nap.edu/catalog/11622

9. Russell RB, Green NS, Steiner CA, Meikle S, Howse JL, Poschman K, et al. Cost of Hospitalization for Preterm and Low Birth Weight Infants in the United States. Pediatrics. 2007 Jul 1;120(1):e1–9.

10. Lim G, Tracey J, Boom N, Karmakar S, Wang J, Berthelot J-M, et al. CIHI survey: Hospital costs for preterm and small-for-gestational age babies in Canada. Healthc Q Tor Ont. 2009;12(4):20–4.

11. Statistics Canada. CANSIM - 102-0562 - Leading causes of death, infants, by sex, Canada [Internet]. 2015 [cited 2016 Apr 4]. Available from: http://www5.statcan.gc.ca/cansim/pick-choisir?lang=eng&p2=33&id=1020562

12. Organization for Economic Cooperation and Development (OECD). Infant mortality rates [Internet]. OECD Publishing; 2014 Jun. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/infant-mortality-rates/indicator/english_83dea506-en

13. Synnes AR, Buchanan L, Ruth C, Albersheim S. Management of the newborn delivered at the threshold of viability. BC Med J. 2008 Nov;50(9):498–508.

14. Stoll BJ, Hansen NI, Bell EF, Shankaran S, Laptook AR, Walsh MC, et al. Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network. Pediatrics. 2010 Sep;126(3):443–56.

15. Volpe JJ. Neurology of the Newborn. 5th ed. Philadelphia, PA: Elsevier; 2008.

16. Tyson JE, Younes N, Verter J, Wright LL. Viability, morbidity, and resource use among newborns of 501- to 800-g birth weight. National Institute of Child Health and Human Development Neonatal Research Network. JAMA. 1996 Nov 27;276(20):1645–51.

17. Singer L, Yamashita T, Lilien L, Collin M, Baley J. A longitudinal study of developmental outcome of infants with bronchopulmonary dysplasia and very low birth weight. Pediatrics. 1997 Dec;100(6):987–93.

18. Madan JC, Kendrick D, Hagadorn JI, Frantz ID. Patent Ductus Arteriosus Therapy: Impact on Neonatal and 18-Month Outcome. Pediatrics. 2009 Feb 1;123(2):674–81.

19. Rees CM, Pierro A, Eaton S. Neurodevelopmental outcomes of neonates with medically and surgically treated necrotizing enterocolitis. Arch Dis Child - Fetal Neonatal Ed. 2007 May 1;92(3):F193–8.

20. Vohr BR, Msall ME, Wilson D, Wright LL, McDonald S, Poole WK. Spectrum of gross motor function in extremely low birth weight children with cerebral palsy at 18 months of age. Pediatrics. 2005 Jul;116(1):123–9.

21. Davis AS, Hintz SR, Goldstein RF, Ambalavanan N, Bann CM, Stoll BJ, et al. Outcomes of extremely preterm infants following severe intracranial hemorrhage. J Perinatol Off J Calif Perinat Assoc. 2014 Mar;34(3):203–8.

22. 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 Aug 1;128(2):e348–57.

23. De Jesus LC, Pappas A, Shankaran S, Li L, Das A, Bell EF, et al. Outcomes of small for gestational age infants born at <27 weeks' gestation. J Pediatr. 2013 Jul;163(1):55–60.e1–3.

24. Casey BM, McIntire DD, Leveno KJ. The Continuing Value of the Apgar Score for the Assessment of Newborn Infants. N Engl J Med. 2001 Feb 15;344(7):467–71.

25. The International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. Lancet. 1993 Jul 24;342(8865):193–8.

26. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. J Pediatr. 2001 Jan;138(1):92–100.

27. Li Y, Yan J, Li M, Xiao Z, Zhu X, Pan J, et al. Addition of SNAP to perinatal risk factors improves the prediction of bronchopulmonary dysplasia or death in critically ill preterm infants. BMC Pediatr. 2013;13:138.

28. Rautonen J, Mäkelä A, Boyd H, Apajasalo M, Pohjavuori M. CRIB and SNAP: assessing the risk of death for preterm neonates. Lancet Lond Engl. 1994 May 21;343(8908):1272–3.

29. Vohr BR, Wright LL, Dusick AM, Mele L, Verter J, Steichen JJ, et al. Neurodevelopmental and Functional Outcomes of Extremely Low Birth Weight Infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993–1994. Pediatrics. 2000 Jun 1;105(6):1216–26.

30. Aylward GP. Neurodevelopmental outcomes of infants born prematurely. J Dev Behav Pediatr JDBP. 2005 Dec;26(6):427–40.

31. Chow SS. Report of the Australia and New Zealand Neonatal Network 2012. Sydney: ANZNN; 2014.

32. Luoma L, Herrgård E, Martikainen A, Ahonen T. Speech and language development of children born at < or = 32 weeks' gestation: a 5-year prospective follow-up study. Dev Med Child Neurol. 1998 Jun;40(6):380–7.

33. Spencer-Smith MM, Spittle AJ, Lee KJ, Doyle LW, Anderson PJ. Bayley-III Cognitive and Language Scales in Preterm Children. Pediatrics. 2015 May;135(5):e1258–65.

34. Limperopoulos C, Bassan H, Gauvreau K, Robertson RL, Sullivan NR, Benson CB, et al. Does Cerebellar Injury in Premature Infants Contribute to the High Prevalence of Long-term Cognitive, Learning, and Behavioral Disability in Survivors? PEDIATRICS. 2007 Sep 1;120(3):584–93.

35. Synnes AR, Anson S, Baum J, Usher L. Incidence and pattern of hearing impairment in children with ≤ 800 g birthweight in British Columbia, Canada: Hearing impairment in ELBW children. Acta Paediatr. 2012 Feb;101(2):e48–54.

36. Wood NS, Marlow N, Costeloe K, Gibson AT, Wilkinson AR. Neurologic and Developmental Disability after Extremely Preterm Birth. N Engl J Med. 2000 Aug 10;343(6):378–84.

37. Spittle A, Orton J, Anderson PJ, Boyd R, Doyle LW. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. Cochrane Database Syst Rev. 2015;11:CD005495.

38. Andrews B, Myers P, Lagatta J, Meadow W. A Comparison of Prenatal and Postnatal Models to Predict Outcomes at the Border of Viability. J Pediatr. 2016 Jun;173:96–100.

39. Younge N, Smith PB, Gustafson KE, Malcolm W, Ashley P, Cotten CM, et al. Improved survival and neurodevelopmental outcomes among extremely premature infants born near the limit of viability. Early Hum Dev. 2016 Apr;95:5–8.

40. Jarjour IT. Neurodevelopmental outcome after extreme prematurity: a review of the literature. Pediatr Neurol. 2015 Feb;52(2):143–52.

41. Vohr BR. How should we report early childhood outcomes of very low birth weight infants? Semin Fetal Neonatal Med. 2007 Oct;12(5):355–62.

42. Rysavy MA, Li L, Bell EF, Das A, Hintz SR, Stoll BJ, et al. Between-hospital variation in treatment and outcomes in extremely preterm infants. N Engl J Med. 2015 May 7;372(19):1801–11.

43. The BOOST-II Australia and United Kingdom Collaborative Groups. Outcomes of Two Trials of Oxygen-Saturation Targets in Preterm Infants. N Engl J Med. 2016 Feb 25;374(8):749–60.

44. Bayley N. Bayley Scales of Infant Development. III. San Antonio, TX: Harcourt Assessment Inc.; 2006.

45. Maccow G. Bayley III: Administration and Scoring [Internet]. Pearson Education; 2008 [cited 2016 May 30]. Available from: http://images.pearsonclinical.com/images/PDF/Bayley-III_Webinar.pdf

46. Chow SS., Le Marsney R, Hossain S, Haslam R, Lui K. Report of the Australia and New Zealand Neonatal Network 2013. Sydney: ANZNN; 2015.

47. 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 Aug;47(8):571–6.

48. A Guide to Cerebral Palsy [Internet]. Cerebral Palsy Association of British Columbia; 2006 [cited 2016 May 30]. Available from:

http://bccerebralpalsy.com/wp-content/uploads/2014/03/a-guide-to-cerebral-palsy.original.pdf

49. Bracewell M, Marlow N. Patterns of motor disability in very preterm children. Ment Retard Dev Disabil Res Rev. 2002;8(4):241–8.

50. Kuban KCK, Allred EN, O'Shea M, Paneth N, Pagano M, Leviton A, et al. An algorithm for identifying and classifying cerebral palsy in young children. J Pediatr. 2008 Oct;153(4):466–72.

51. 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 Apr 1;39(4):214–23.

52. Serenius F, Källén K, Blennow M, et al. Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in sweden. JAMA. 2013 May 1;309(17):1810–20.

53. Fierson WM, Saunders RA, Good W, Palmer EA, Phelps D, Reynolds J, et al. Screening Examination of Premature Infants for Retinopathy of Prematurity. Pediatrics. 2013 Jan 1;131(1):189–95.

54. Schmidt B, Whyte RK, Asztalos EV, Moddemann D, Poets C, Rabi Y, et al. Effects of targeting higher vs lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. JAMA. 2013 May 22;309(20):2111–20.

55. Synnes A, Luu TM, Moddemann D, Church P, Vincer M, Ballantyne M, et al. Determinants of Developmental Outcomes in a Very Preterm Canadian Cohort. Accept Arch Child. 2016 Sep 9;

56. Cristobal R, Oghalai JS. Hearing loss in children with very low birth weight: current review of epidemiology and pathophysiology. Arch Dis Child Fetal Neonatal Ed. 2008 Nov;93(6):F462–8.

57. Canadian Neonatal Follow-Up Network. 18-Month Corrected Age Assessment Manual, Version 5 [Internet]. Canadian Neonatal Follow-Up Network; 2012 [cited 2016 Jul 29]. Available from: http://www.cnfun.ca/LinkClick.aspx?fileticket=d4p7mZoXWDU%3d&tabid=68

58. Marlow N, Abbott J, Field D, Johnson S, Huertas A, Jones H, et al. Report of a BAPM/RCPCH Working Group. Classification of health status at 2 years as a perinatal outcome. Version 1.0. British Association of Perinatal Medicine; 2008 Jan.

59. Doyle LW, Roberts G, Anderson PJ. Outcomes at Age 2 Years of Infants < 28 Weeks' Gestational Age Born in Victoria in 2005. J Pediatr. 2010 Jan;156(1):49–53.e1.

60. Synnes AR, Anson S, Arkesteijn A, Butt A, Grunau RE, Rogers M, et al. School entry age outcomes for infants with birth weight ≤ 800 grams. J Pediatr. 2010 Dec;157(6):989–94.e1.

61. Doyle LW, Anderson PJ, Haslam R, Lee KJ, Crowther C, Australasian Collaborative Trial of Magnesium Sulphate (ACTOMgSO4) Study Group. School-age outcomes of very preterm infants after antenatal treatment with magnesium sulfate vs placebo. JAMA. 2014 Sep 17;312(11):1105–13.

62. Moore GP, Lemyre B, Barrowman N, Daboval T. Neurodevelopmental outcomes at 4 to 8 years of children born at 22 to 25 weeks' gestational age: a meta-analysis. JAMA Pediatr. 2013 Oct;167(10):967–74.

63. Holsti L, Grunau RVE, Whitfield MF. Developmental coordination disorder in extremely low birth weight children at nine years. J Dev Behav Pediatr JDBP. 2002 Feb;23(1):9–15.

64. Anderson PJ, Doyle LW, Victorian Infant Collaborative Study Group. Executive functioning in school-aged children who were born very preterm or with extremely low birth weight in the 1990s. Pediatrics. 2004 Jul;114(1):50–7.

65. Nosarti C, Reichenberg A, Murray RM, et al. Preterm birth and psychiatric disorders in young adult life. Arch Gen Psychiatry. 2012 Jun 1;69(6):610–7.

66. Allin M, Matsumoto H, Santhouse AM, Nosarti C, AlAsady MH, Stewart AL, et al. Cognitive and motor function and the size of the cerebellum in adolescents born very pre-term. Brain J Neurol. 2001 Jan;124(Pt 1):60–6.

67. Moster D, Lie RT, Markestad T. Long-Term Medical and Social Consequences of Preterm Birth. N Engl J Med. 2008 Jul 17;359(3):262–73.

68. Breeman LD, Jaekel J, Baumann N, Bartmann P, Wolke D. Preterm Cognitive Function Into Adulthood. Pediatrics. 2015 Sep;136(3):415–23.

69. Eryigit Madzwamuse S, Baumann N, Jaekel J, Bartmann P, Wolke D. Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. J Child Psychol Psychiatry. 2015 Aug;56(8):857–64.

70. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet. 2008 Jan 19;371(9608):261–9.

71. MacDonald H, American Academy of Pediatrics. Committee on Fetus and Newborn. Perinatal care at the threshold of viability. Pediatrics. 2002 Nov;110(5):1024–7.

72. Peerzada JM, Richardson DK, Burns JP. Delivery room decision-making at the threshold of viability. J Pediatr. 2004 Oct;145(4):492–8.

73. Higgins RD, Delivoria-Papadopoulos M, Raju TNK. Executive Summary of the Workshop on the Border of Viability. Pediatrics. 2005 May 1;115(5):1392–6.

74. Partridge JC, Freeman H, Weiss E, Martinez AM. Delivery room resuscitation decisions for extremely low birthweight infants in California. J Perinatol Off J Calif Perinat Assoc. 2001 Feb;21(1):27–33.

75. Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive Care for Extreme Prematurity — Moving beyond Gestational Age. N Engl J Med. 2008 Apr 17;358(16):1672–81.

76. Schmidt B, Roberts RS, Davis P, Doyle LW, Barrington KJ, Ohlsson A, et al. Long-Term Effects of Caffeine Therapy for Apnea of Prematurity. N Engl J Med. 2007 Nov 8;357(19):1893–902.

77. Williams J, Lee KJ, Anderson PJ. Prevalence of motor-skill impairment in preterm children who do not develop cerebral palsy: a systematic review. Dev Med Child Neurol. 2010 Mar;52(3):232–7.

78. CNFUN | Canadian Neonatal Follow-Up Network | RSCN [Internet]. [cited 2016 Aug 22]. Available from: http://www.cnfun.ca/

79. The MICare Project [Internet]. [cited 2016 Aug 22]. Available from: http://www.cnfun.ca/TheMICareProject.aspx

80. Hosmer DW, Lemeshow S. Applied Logistic Regression: Hosmer/Applied Logistic Regression [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2000 [cited 2016 Aug 22]. Available from: http://doi.wiley.com/10.1002/0471722146

81. Kusuda S, Fujimura M, Uchiyama A, Nakanishi H, Totsu S, for Neonatal Research Network, Japan. Identification of practices and morbidities affecting the mortality of very low birth weight infants using a multilevel logistic analysis: clinical trial or standardisation? BMJ Open. 2013;3(8):e003317.

82. Kusuda S, Fujimura M, Sakuma I, Aotani H, Kabe K, Itani Y, et al. Morbidity and mortality of infants with very low birth weight in Japan: center variation. Pediatrics. 2006 Oct;118(4):e1130–8.

83. Fily A, Pierrat V, Delporte V, Breart G, Truffert P, EPIPAGE Nord-Pas-de-Calais Study Group. Factors associated with neurodevelopmental outcome at 2 years after very preterm birth: the population-based Nord-Pas-de-Calais EPIPAGE cohort. Pediatrics. 2006 Feb;117(2):357–66.

84. O'Shea TM, Allred EN, Kuban KCK, Hirtz D, Specter B, Durfee S, et al. Intraventricular Hemorrhage and Developmental Outcomes at 24 months of age in Extremely Preterm Infants. J Child Neurol. 2012 Jan;27(1):22–9.

85. Adams M, Borradori-Tolsa C, Bickle Graz M, Grunt S, Weber P, Capone Mori A, et al. Recommendations of the Swiss Society of Neonatology, the Swiss Society of Developmental Pediatrics and the Swiss Society of Neuropediatrics: Follow-up assessment of high-risk newborns in Switzerland. Paediatrica. 2014;25(5):6–10.

86. Serenius F, Blennow M, Maršál K, Sjörs G, Källen K, EXPRESS Study Group. Intensity of Perinatal Care for Extremely Preterm Infants: Outcomes at 2.5 Years. Pediatrics. 2015 Apr 20;

87. Smylie J, Fell D, Ohlsson A, Joint Working Group on First Nations Indian Inuit, Métis Infant Mortality of the Canadian Perinatal Surveillance System. A review of Aboriginal infant mortality rates in Canada: striking and persistent Aboriginal/non-Aboriginal inequities. Can J Public Health Rev Can Santé Publique. 2010 Apr;101(2):143–8.

88. O'Driscoll T, Kelly L, Payne L, St Pierre-Hansen N, Cromarty H, Minty B, et al. Delivering away from home: the perinatal experiences of First Nations women in northwestern Ontario. Can J Rural Med Off J Soc Rural Physicians Can J Can Médecine Rurale J Off Société Médecine Rurale Can. 2011;16(4):126–30.

89. Riddell CA, Hutcheon JA, Dahlgren LS. Differences in obstetric care among nulliparous First Nations and non–First Nations women in British Columbia, Canada. CMAJ Can Med Assoc J. 2016 Feb 2;188(2):E36–43.

90. Lisonkova S, Haslam M, Dahlgren L, Chen I, Synnes A, Lim K. Maternal morbidity and perinatal outcomes among Women in rural versus urban areas. Can Med Assoc J [Internet]. 2016 Sep 26 [cited 2016 Aug 22]; Available from: http://www.cmaj.ca/cgi/doi/10.1503/cmaj.151382

91. Findlay L, Kohen D, Miller A. Developmental milestones among Aboriginal children in Canada. Paediatr Child Health. 2014 May;19(5):241–6.

92. Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW, Victorian Infant Collaborative Group. Underestimation of developmental delay by the new Bayley-III Scale. Arch Pediatr Adolesc Med. 2010 Apr;164(4):352–6.

93. Flynn JR. Massive IQ gains in 14 nations: What IQ tests really measure. Psychol Bull. 1987;101(2):171–91.

94. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. The Bayley-III cognitive and language scales: how do scores relate to the Bayley-II? Arch Dis Child. 2011 Apr 1;96(Suppl 1):A39–40.

95. Askie LM, Brocklehurst P, Darlow BA, Finer N, Schmidt B, Tarnow-Mordi W. NeOProM: Neonatal Oxygenation Prospective Meta-analysis Collaboration study protocol. BMC Pediatr. 2011 Jan 17;11:6.

96. Vanderveen JA, Bassler D, Robertson CMT, Kirpalani H. Early interventions involving parents to improve neurodevelopmental outcomes of premature infants: a meta-analysis. J Perinatol. 2009 Jan 15;29(5):343–51.

97. Bode MM, D'Eugenio DB, Mettelman BB, Gross SJ. Predictive validity of the Bayley, Third Edition at 2 years for intelligence quotient at 4 years in preterm infants. J Dev Behav Pediatr. 2014 Dec;35(9):570–5.

98. Spittle AJ, Spencer-Smith MM, Eeles AL, Lee KJ, Lorefice LE, Anderson PJ, et al. Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? Dev Med Child Neurol. 2013 May;55(5):448–52.

99. Luttikhuizen dos Santos ES, de Kieviet JF, Königs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. Early Hum Dev. 2013 Jul;89(7):487–96.

100. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA. 2003 May 21;289(19):2554–9.

101. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. Am Stat. 2016 Apr 2;70(2):129–33.

102. Synnes AR, Macnab YC, Qiu Z, Ohlsson A, Gustafson P, Dean CB, et al. Neonatal intensive care unit characteristics affect the incidence of severe intraventricular hemorrhage. Med Care. 2006 Aug;44(8):754–9.

103. Palisano R, Rosenbaum P, Bartlett D, Livingston M. Gross Motor Function Classification System – Expanded and Revised [Internet]. CanChild Centre for Childhood Disability Research, McMaster University; 2007 [cited 2016 Sep 30]. Available from: https://www.cpqcc.org/sites/default/files/documents/HRIF_QCI_Docs/GMFCS-ER.pdf

Appendix

Figure 2: Risk factors for neurodevelopmental impairment between 18-24 months of age

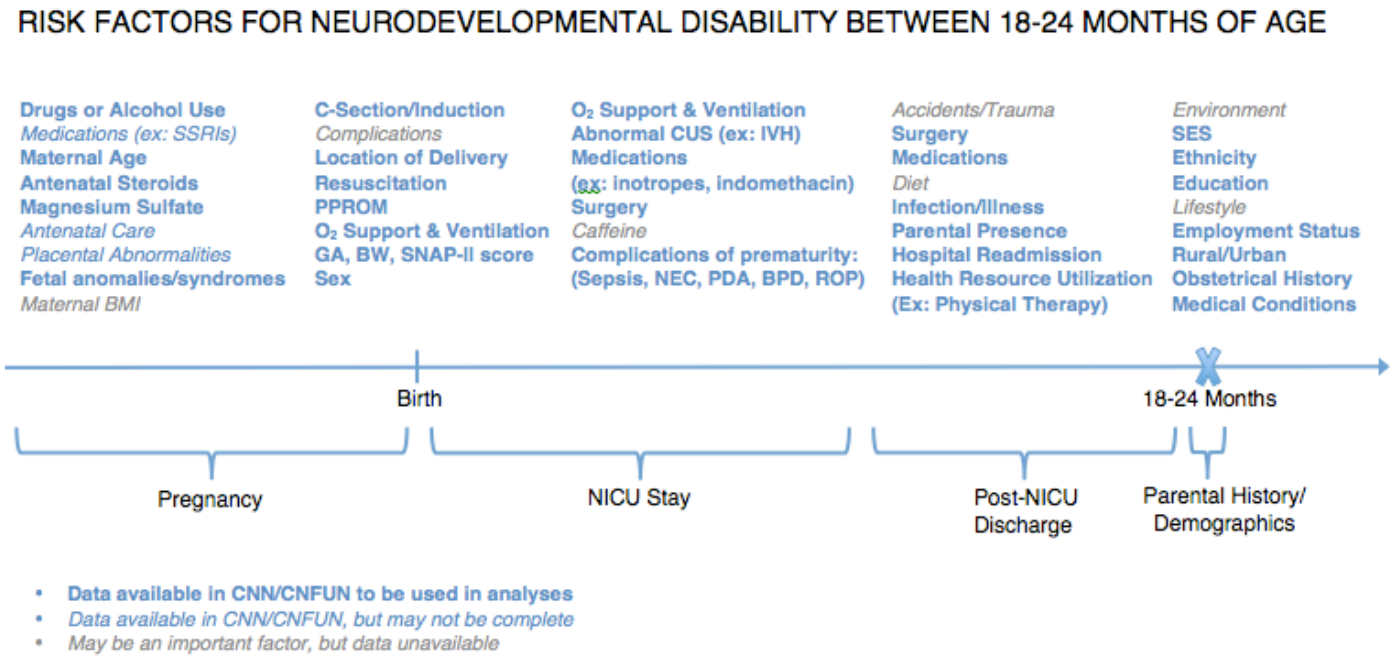


Table 15: Gross Motor Function Classification System (GMFCS) Functional Ability Classification Levels at 18-24 months (51,103)

GMFCS Level	Description of Functional Ability
1	• Move in and out of sitting and floor sit with both hands free to manipulate objects • Crawl on hands and knees, pull to stand and take steps holding onto furniture • Can walk without the need for any assistive mobility device
2	• Maintain floor sitting but may need to use hands for support to maintain balance • Creep on stomach or crawl on hands and knees • May pull to stand and take steps holding onto furniture
3	• Maintain floor sitting when lower back is supported • Roll and creep forward on stomach
4	• Have head control but trunk support is required for floor sitting • Can roll to supine and may roll to prone
5	• Physical impairments limit voluntary control of movement • Unable to maintain antigravity head and neck postures in prone and sitting • Require adult assistance to roll

Table 16: Literature and internet search inclusion criteria for national or regional neonatal research networks

	Neonatal Research Network Inclusion Criteria
1	Admit or recruit very premature infants and collect data on this population
2	>1 research site systematically following-up very preterm infants and investigators from >1 site involved with the network
3	Published >1 research protocol in scientific literature in English language within the past 10 years
4	Have clearly defined follow-up information on neurodevelopmental outcome (including moderate or severe neurodevelopmental impairment components) at the 18-24 month assessment

Table 17: Description of variables used in study

Variable	Type	Values/Levels/Definition
Socio-demographic		
Ethnicity	Categorical	First Nations East Asian South Asian White Other/unknown
Education	Categorical	<High school Completed high school Post-secondary
Employment status	Categorical	Unemployed Student or homemaker Employed part- / full-time Other/unknown
Lone parent	Binary	Yes, No
Pregnancy		
Maternal age	Nominal Categorical	Value between 14-70 <25 25-34 35+
Parity	Binary	Primiparous: Yes, No
Diabetes	Binary	Diabetes Mellitus (gestational or chronic): Yes, No
Pre-eclampsia	Binary	Hypertension (BP>140/90 twice, >4 hours apart) and proteinuria: Yes, No
Antenatal steroids	Binary	Antenatal exposure to Dexamethasone or Betamethasone: Yes, No
Magnesium sulfate	Binary	Antenatal exposure to magnesium sulfate: Yes, No
Drugs	Binary	Antenatal marijuana, cocaine, alcohol, heroin, or other illicit drug exposure: Yes, No
Cigarettes	Binary	Antenatal exposure to firsthand cigarette use: Yes, No
Non-singleton	Binary	Twins or multiple gestation pregnancy: Yes, No

Delivery and Birth		
Apgar score 5min	Nominal	Value between 0–15
Apgar5<7	Binary	Apgar 5 min < 7: Yes, No
Gestational age (GA)	Nominal	Value between 22 – 28^{6} weeks
Birth weight	Nominal	Value between 350g – 2500g
Small for gestational age (SGA)	Binary	Weight below 10^{th} percentile: Yes, No
Sex	Binary	Male: Yes, No
Clinical chorioamnionitis	Binary	Confirmed chorioamnionitis: Yes, No
Premature rupture of membranes (PROM)	Binary	PROM: Yes, No
Mode of delivery	Binary	Vaginal, Caesarean
Presentation:	Binary	Vertex: Yes, No
	Binary	Breech: Yes, No
	Binary	Transverse: Yes, No
	Binary	Other: Yes, No
Outborn	Binary	Delivery outside hospital setting: Yes, No
NICU		
SNAP-II score	Nominal	Values between 0-40
SNAP-II score>20	Binary	SNAP-II score >20: Yes, No
Intraventricular hemorrhage (IVH)- all grades	Ordinal	Grade: None, 1, 2, 3, 4
IVH – severe	Binary	IVH grade 3-4: Yes, No
Sepsis	Binary	Positive blood/CSF; Early onset: Yes, No Late onset: Yes, No
Bronchopulmonary dysplasia (BPD)	Binary	Requires oxygen or respiratory support at 36 weeks corrected PMA: Yes, No
Retinopathy of prematurity (ROP) – all	Ordinal	Grade: None, 1, 2, 3, 4
ROP – severe	Binary	ROP ≥ grade 3: Yes, No
Necrotising enterocolitis (NEC)	Binary	Incidence of Bell's ≥ grade 2 NEC: Yes, No
Patent ductus arteriosis (PDA)	Binary	Medically or surgically treated PDA: Yes, No