INVESTIGATION OF THE TEST CHARACTERISTICS OF TWO SCREENING TOOLS IN COMPARISON TO A GOLD STANDARD ASSESSMENT TO DETECT DEVELOPMENTAL DELAY: A PILOT STUDY

by

Lisa Marie Currie

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2011

DALHOUSIE UNIVERSITY

DEPARTMENT OF COMMUNITY HEALTH AND EPIDEMIOLOGY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled, "INVESTIGATION OF THE TEST CHARACTERISTICS OF TWO SCREENING TOOLS IN COMPARISON TO A GOLD STANDARD ASSESSMENT TO DETECT DEVELOPMENTAL DELAY: A PILOT STUDY" by Lisa Marie Currie in partial fulfilment of the requirements for the degree of Master of Science.

Dated: 19 August 2011

Supervisor: _____

Readers: _____

_____

_____

DALHOUSIE UNIVERSITY

DATE: 19 August 2011

AUTHOR: Lisa Marie Currie

TITLE: INVESTIGATION OF THE TEST CHARACTERISTICS OF TWO SCREENING TOOLS IN COMPARISON TO A GOLD STANDARD ASSESSMENT TO DETECT DEVELOPMENTAL DELAY: A PILOT STUDY

DEPARTMENT OR SCHOOL: Department of Community Health and Epidemiology

DEGREE: MSc CONVOCATION: October YEAR: 2011

_____

Signature of Author

TABLE OF CONTENTS

LIST OF TABLES

Abstract

There is minimal information available regarding test characteristics of the Rourke and the NDDS, two tools commonly used to screen for developmental delay. The objectives are to (a) generate preliminary descriptive data about the population and outcomes of interest, (b) determine test characteristics of the tools compared to the gold standard assessment, BSITD-III. Thirty-six month old children at high risk of developmental delay were recruited from the Perinatal Follow-up Program at the IWK Health Centre in Halifax, NS. The Rourke and NDDS results were obtained via parental report, the BSITD-III via clinical assessment. Results suggest that both tools may possess appropriate test characteristics to screen for developmental delay. Both perform more accurately when the criterion for delay is extended to two flagged areas of concern on the assessment tools. In conclusion, both tools appear to be sensitive to detecting developmental delay. Further investigation via a full scale study is warranted.

LIST OF ABBREVIATIONS AND SYMBOLS USED

Bayley Scales of Infant and Toddler Development-III	**BSITD-III**
Bayley Scales of Infant Development version II	BSID-II
Positive Screen for Developmental Delay on BSITD-III	BSITD-III +
Negative Screen for Developmental Delay on BSITD-III	BSITD-III -
Denver Developmental Screening Tool	DDST
False Negative	FN
False Positive	FP
Izaak Walton Killam Health Centre	IWK Health Centre
Negative Predictive Value	NPV
Nipissing District Developmental Screen	NDDS
Flag indicating area of concern for further assessment of developmental delay on NDDS	NDDS+
No areas of concern noted on NDDS	NDDS-
Odds Ratio	OR
Rourke Baby Record	Rourke
Flag indicating area of concern for further assessment of developmental delay on Rourke	Rourke +
No areas of concern noted on Rourke	Rourke -
Positive Predictive Value	PPV
True Positive	TP
True Negative	TN
Standard Deviation	SD
95% Confidence Interval	95% CI

ACKNOWLEDGEMENTS

I foremost wish to acknowledge the contributions of my thesis supervisor, Dr. Linda Dodds, for her dedication towards the advancement of my knowledge in perinatal epidemiology. I also wish to acknowledge my committee, Dr. Gordon Flowerdew, Dr. Sarah Shea and Dr. Michael Vincer. Their contributions assisted in the comprehensiveness of both the project and my academic development. Further, to Dr. Jennifer McLean and Dr. Robin Walker for their role in the project's development.

I would like to extend my gratitude to the staff of the Perinatal Follow-up Program for their assistance and dedication towards the project's recruitment and data collection procedures.

I wish to acknowledge my family for supporting my decision to pursue this training, and in providing reassurance and encouragement throughout the entire process.

Finally, I gratefully acknowledge the families who agreed to participate in this study. Their willingness to provide information regarding their child's development demonstrated true altruism towards the development of knowledge in this area of research. Without their cooperation, this project would not have been possible.

CHAPTER 1: INTRODUCTION

The proposed research topic was designed to assess the test characteristics of the Rourke Baby Record (Rourke) and Nipissing District Developmental Screen (NDDS) in detection of delayed development, in comparison to the gold standard assessment, the Bayley Scales of Infant and Toddler Development III (BSITD-III). Physicians and other health care professionals currently use developmental screening tools to determine if a child is at risk of developmental delay. If a child scores positive on the screen, they are typically referred for developmental assessment which is often done using the BSITD-III. The Rourke is the most frequently used surveillance tool in Nova Scotia and to a lesser degree the NDDS. In Ontario, the NDDS is recommended as an additional screening tool to the Rourke (Rourke, 2009). However, neither of these tools has been evaluated to determine its ability to predict developmental delay (Personal Communication, 2009).

This pilot observational study assessed high risk children with the Rourke, NDDS and BSITD-III to determine if results on the Rourke and NDDS predict performance on the BSITD-III. The cohort consisted of 31 thirty-six month old children born at the IWK Health Centre. Criteria for participants to be considered high risk were (a) a gestational age of $\leq$31 weeks or (b) $\leq$ 1500 grams at birth, or (c) neurological injury at or immediately following birth, as infants with any of these characteristics are at risk of developmental delay. Recruitment was based on children enrolled in the Perinatal Follow-up Program at the IWK Health Centre and scheduled for a visit during the study period.

Participants' parents were contacted by telephone approximately one week prior to administration of the BSITD-III. In the telephone interview, parents were asked

questions from the developmental section of the Rourke Baby Record: thirty-six months. The NDDS thirty-six months was completed one week later at the Perinatal Follow-Up Clinic scheduled appointment. The BSITD-III, a direct developmental assessment of child development, was also completed at this visit.

The purpose of this pilot study was to ultimately test procedures to determine appropriate methodology and recruitment procedures for execution of a full observational study to assess the test characteristics of these two tools used to screen for developmental delay. Further, this study aimed to determine (a) the sensitivity and specificity of the Rourke and the NDDS each to screen for developmental delay in high risk groups, and (b) the rate of false positives and false negatives emerging from use of the Rourke and NDDS, compared to the gold standard assessment, BSITD-III.

This research provides insight to the capacities of the Rourke and the NDDS in high risk children by comparing results of these tools with those obtained using the gold standard test, the BSITD-III. The BSITD-III's predictive ability to detect developmental delay has been questioned, especially in high risk infants (Anderson et al., 2010). However, it remains the "gold standard" for assessment of development in infants and toddlers (Bayley, 2006).

CHAPTER 2: LITERATURE REVIEW

2.1 Population at Risk

Developmental delay refers to a limitation in gross motor, fine motor, cognitive, language, or personal-social skills below age expected performance (Koseck, 1999). Statistics Canada data indicates that 1.6% of all children aged 0-4 years have some form of developmental delay, such as vision, auditory or mobility limitations or learning disability (Cosette and Duclos, 2001). This rises to approximately 17% by eighteen years of age (Boyle, Decoufle and Yeargin-Allsopp, 2004). Rates are as high as 23% for those born extremely prematurely (Lorenz et al., 1998).

Early detection of delay can result in earlier intervention and implementation of resources that minimize the functional impact of delay. Leib, Benfield and Guidubaldi (1980) indicated that an early intervention and stimulation program is integral for high risk infants with potential developmental delay. If developmental delay can be accurately identified early in a child's development, it is more likely that, with early intervention, the child will develop skills necessary to optimize the outcome (Leib et al., 1980). Several medical states can lead to the classification of high risk, including prematurity, low birth weight and neurological injury (Koseck and Harris, 2004).

Goyen and Lui (2002) indicated that some preterm infants who otherwise appear without delay may have marked fine and gross motor developmental delay upon formal assessment, suggesting that delay may not be evident without assessment. Further, Janssen, Nijhuis-van der Sanden, Akkermans, Oostendorp and Kollée et al (2008) suggested that preterm infants were at an increased risk of motor delay and demonstrated

suboptimal behaviours in test taking at age two to three years. Early assessment of these infants can lead to early access to care services to decrease functional impact of developmental delay.

Kono, Mishina, Sato, Watanabe and Honma (2008) suggest that infants born with very low birth weights (<1500 grams) had greater developmental delay in fine motor skills, following commands and communication skills, as well as potential differences in behaviour at corrected eighteen months of age in comparison to those of normal birth weight.

Infants who experience an adverse neurological event, such as birth asphyxia, cerebrovascular accident or haemorrhage often experience developmental delay associated with brain damage as well as acute illness early in life. More extensive forms of periventricular haemorrhage, typically associated with preterm birth, have been found to have a severe impact on function across all domains, as measured by the mental scale and psychomotor scales of the Bayley Scales of Infant Development -Version I, at two years of age (Catto-Smith, Yu, Bajuk, Orgill, Astbury, 1985)

2.2 Assessment

Family physicians, paediatricians and/or neonatologists typically have regular contact with high risk infants for medical follow-up and are often the health care providers involved with the identification of developmental delay. It is imperative that a tool is readily available that is not only time efficient for a medical practice, but also sensitive to detect deficits in development. The BSITD-III is a well documented assessment tool, but is costly and lengthy to administer, hence not an ideal choice for

physicians. The Rourke is a relatively recent, comprehensive, evidence based, integrated primary care practice tool (Rourke et al., 2009). However, the properties to identify developmental concern in infants of the Rourke are largely undetermined. The NDDS is similar to the Rourke in that it is widely used but has minimal literature to support its ability to detect delayed development (Dahinten and Ford, 2004, NDDS, 2007).

In the past, the DDST was a tool used by many physicians. The Canadian Task Force on the Periodic Health Examination excluded the DDST in 1994 as it was resulting in over referral due to an increased number of false positive identifications of developmental delay (Limbos and Joyce, 1999). Instead of administering the DDST, physicians now frequently rely on the Rourke or the NDDS to identify delay (Limbos et al., 1999).

Parental report, as obtained with both the Rourke and the NDDS, is suggested to provide reliable information about children's development. However, there is the risk that parental report may be confounded by social determinants, such as literacy rate and performance expectations. Hence, parental report alone is not directly predictive of child developmental performance (Alward, 2009). As standardized assessments are often used to determine a child's referral to additional health care services such as pediatric medical specialities or allied health professionals, the more sensitive an assessment is to assessing differences in a child's behaviour, the greater likelihood of detecting developmental delay (Connelly et al., 2006).

2.3 BAYLEY SCALES OF INFANT AND TODDLER DEVELOPMENT VERSION III

The BSITD-III was released in 2006 and differs from the BSID-II in that it is more user-friendly for both the administrator and the child, has new normative values from the 2000 Census, has simplified scoring and encourages parental contribution (Pearson Assessment and Administration, 2009). It has three scales: cognitive, motor (subdivided to gross and fine) and language (subdivided to receptive and expressive) (Bayley, 2006). The assessment is more comprehensive than its predecessor, as the BSID-II consisted of only a mental and a motor scale (Bayley, 2006). Content was revised, and several components were modified or omitted. There are two parental questionnaires pertaining to social emotional and adaptive behaviour scales (Bayley, 2006). The authors acknowledged and attempted to overcome limitations to the psychometric properties of the BSID-II in the development of the BSITD-III (Bayley, 2006). The tool was standardized on a sample of 1700 healthy children. Children at high risk of developmental delay due to birth complications were excluded from the standardization process. Due to the novelty of this tool, there is limited published evidence in the literature pertaining to the psychometric properties of the BSITD-III. However, the BSITD-III remains the gold standard assessment of choice for detection of developmental delay for children at 36 months of age (Bayley, 2006).

2.4 Measurement criteria of BSITD-III

Previous studies have indicated that assessment tools utilized in the clinical environment should have a reliability correlation coefficient of r= 0.9 in comparison to the gold standard assessment if used for diagnostic purposes, or r=0.8 if used as a screening tool (Bracken, 1987).

The authors of the BSITD-III assessed the internal consistency of the test (Bayley, 2006). It was found that the reliability coefficients of the subtests averaged r= 0.95 for thirty-six month old children across all domains (Bayley, 2006). In assessment of thirty-six month children at elevated risk of developmental delay, internal consistency was again high for all subscales (r=0.98) (Bayley, 2006).

The analysis of test-retest reliability suggests a high level of reliability when the tests were administered within an averaged six day interval. Specifically, for children age 33-42 months, the test-retest reliability were as follows: cognitive: r=0.86, receptive communication: r=0.90, expressive communication: r= 0.94, fine motor: r= 0.88, gross motor: r=0.83 language composite: r=0.94 motor composite: 0.88 (Bayley, 2006). The overall corrected inter-rater reliability of the BSITD-III is r=0.82 (Bayley, 2006).

The authors assessed the psychometric properties of the BSITD-III of children at high risk of developmental delay in comparison to those at low risk. The results suggest that high risk children scored worse on all subscales and composites than the low risk children, with moderate to large effect sizes for all tests: cognitive ($p<.01$), receptive communication ($p<.01$), expressive communication ($p<.01$), fine motor ($p<.01$), gross motor ($p<.01$), social emotional ($p<.05$), language composite ($p<.01$), motor composite ($p<.01$), suggestive that the test is sensitive to detect developmental delay in at risk children (Bayley, 2006).

The BSITD-III was predictive of developmental delay upon comparison of various at-risk groups to a control group. It detected delay across all subtests for children with Down syndrome ($p<.01$), pervasive developmental disorder ($p<.01$), cerebral palsy

($p<.01$) and specific language impairment ($p<.01$). The BSITD-III was predictive of developmental delay for children with asphyxiation at birth for the cognitive, receptive language, fine motor, gross motor and social emotional subtests ($p<.05$) and was sensitive to detect delay in children born premature in the fine motor subtest only ($p<.05$) (Bayley, 2006).

2.5 The Rourke Baby Record

Studies have indicated that the Rourke is widely used by physicians, specifically to assess healthy child development, recording visits, identifying health issues and high risk concerns (Rourke, Godwin, Rourke, Pearce and Bean, 2009). It is endorsed by the College of Family Physicians of Canada and the Canadian Pediatric Society (Canadian Pediatric Society, 2008) and chart audits suggest that physicians frequently incorporate the Rourke results into their patient charts (Rourke et al., 2009).

Revisions to the Rourke included age appropriate evaluations in accordance with the Canadian Immunization Record as a marker of Healthy Baby Visits (Panagiotou, Rourke, Rourke, Wakefield and Winfield, part 1, 1998). The Rourke acknowledges the importance of assessment of infant development over the first two years of life and classifies high risk infants based upon their environmental situation, such as lower socioeconomic status or abuse (Panagiotou et al., part 2, 1998), but does not account for risk associated with biological factors. The Rourke was revised in 2009 to improve the comprehensiveness of the assessment.

The Rourke instructions for use indicate that it is based on the NDDS (Rourke, 2007). If one or more items are marked as an unmet expectation/of concern on the

questioning, this indicates a need for further assessment (Rourke, 2007). Refer to Appendix A for a review of Rourke: Thirty-Six Month assessment. While the Rourke is comprehensive in its ability to sample through questioning various aspects of child development, the authors report that its purpose is aimed at developmental surveillance, not as a developmental screen.

A developmental surveillance instrument is defined as a "flexible, continuous process whereby knowledgeable health care professionals identify children who may bave developmental problems" ("Identifying Infants", 2006). Surveillance measures may include parental input, observational assessment, reference to medical history and discussions with other health care professionals regarding the child's development. This differs from a developmental screening test, which is defined as a "brief assessment procedure designed to identify children who should receive more intensive diagnosis or assessment" ("Developmental Surveillance", 2001). It is often in the form of a limited standardized assessment. The aim of the screening tool is to determine if further evaluation is necessary. A developmental screen only assesses one point in time in the child's development. The American Academy of Pediatrics Committee on Children with Disabilities recommends that paediatricians continue to complete developmental surveillance, but to include developmental screening tools to improve accuracy of their assessment ("Identifying Infants", 2006).

2.6 Measurement Criteria of the Rourke

Dinkevich, Hupert and Moyer (2001) indicate that the Rourke has yet to be evaluated on clinical based outcomes. Despite the Rourke being widely used, there is

minimal data surrounding the surveillance properties of this tool. No studies could be located that directly assessed the psychometric properties of the Rourke.

2.7 Nipissing District Developmental Screen

The NDDS was created by a committee in Nipissing, Ontario in 1993 to screen for developmental delay. The original goal was to create a screen for children over the age of three. However, it has since been revised to form assessments for children as young as one month of age. The NDDS has been adopted in various locations across Canada, and has been identified as the pediatric screening tool of choice in the Northwest Territories and New Brunswick (NDDS, history, 2007). A nominal cost is associated with use of the NDDS. However, when it was available free of charge to physicians in Ontario, use of the NDDS approximately tripled, thereby highlighting the complexity of factors that lead to tool use (Limbos et al., 2010). A benefit of the NDDS is that it includes a list of recommendations for parents pertaining to a child's specific age (NDDS, 2007). However, minimal literature is available on this screening tool, and information pertaining to the psychometric properties is largely unknown. Refer to Appendix A for an outline of variables assessed on the NDDS: Thirty-Six Month screen.

2.8 Measurement Criteria of the NDDS

Upon assessment of concurrent validity, research suggests that the NDDS is effective in detecting major developmental delay in children but less sensitive to detecting mild developmental delay. However, researchers indicate that coupled with the clinical reasoning skills of a trained health care professional in paediatrics, the validity of the assessment would increase (Dahinten and Ford, 2004).

In terms of detection of delay with one abnormal, identified as a '*no*' response, the NDDS was found to have 56.3% sensitivity and 70.6% specificity in comparison to a one standard deviation of the mean cut-off for the BSID-II to detect delay from a normal population. Sensitivity decreased when detection of delay was defined as two 'no' responses to 43.8%, but specificity increased to 86.3%, compared to a one standard deviation of the mean cut-off for the BSID-II (Dahinten and Ford, 2004). Further, utilizing the one 'no' response, the NDDS had 25.4% over referral (i.e., false positive rate) and 5.9% under referral (i.e., false negative rate) for further assessment compared to the BSID-II (Dahinten and Ford, 2004).

CHAPTER 3: RESEARCH OBJECTIVES

The primary objective of this pilot study was to test participant recruitment procedures and methodology to determine the feasibility of completion of a full scale, adequately powered observational study to determine the test characteristics of the Rourke and NDDS. Secondly, this project aimed to determine the test characteristics of the Rourke and the NDDS compared to the BSITD-III in the assessment of delay in high risk children; specifically, the sensitivity and specificity of the two tests as determined by examining the results compared to the BSITD-III. The final objective of this study was to quantify this association adjusting for demographic characteristics.

CHAPTER 4: METHODOLOGY

4.1 SUMMARY

This pilot observational study was designed to evaluate the Rourke and NDDS against the BSITD-III in detection of developmental delay. High risk thirty-six month old children were assessed using these three developmental tools (n=31). The Rourke and NDDS information were obtained via parental report; the BSITD-III was administered according to the standardized protocol via clinical assessment (Bayley, 2006, administrative manual). Measures of the properties of the screening and surveillance tools and logistic regression were used to analyze the data. Research ethics approval was obtained from the IWK Research Ethics Board (#4998 01451).

4.2 SAMPLE SELECTION

High risk children enrolled in the Perinatal Follow-up Program at the IWK Health Centre and scheduled for a thirty-six month follow-up visit were recruited for this study. Children were primarily from Halifax County, but also included children seen at the Perinatal Follow-up Program travelling clinics in Prince Edward Island, Yarmouth, Antigonish and Sydney, Nova Scotia. Criteria for participants to be considered high risk were: (a) a gestation age of ≤31 weeks or (b) ≤1500 grams at birth, or (c) neurological injury at or immediately following birth. The literature suggests that these children are at increased risk of developmental delay based on their medical condition (Koseck and Harris, 2004).

Exclusion criteria for participation in the study include non-English speaking children, due to the potential inconsistencies with translation of the document, as well as

children with major sensory or physical impairment, as they would not be able to complete the BSITD-III.

4.3 Score Criteria

The cut-off values to indicate need for further assessment for developmental delay on both the Rourke and the NDDS were defined as one or more items noted as an area of concern on the assessment form, as suggested by the administration guidelines. The number of 'no' responses was quantified in this analysis as one 'no' response will serve as the threshold value to suggest further investigation. Two 'no' responses were also assessed. A 'no' response, indicating that the child does not demonstrate the skill, will be termed a 'flag' from herein, to maintain consistency with previous literature (Dahinten and Ford, 2004). The BSITD-III was administered regardless of the results of the Rourke and the NDDS and used to determine the sensitivity and specificity of the assessments to detect developmental delay. A score of ≤ 85 (one standard deviation below the mean) in any domain on the BSITD-III served as the cut-off for normal performance, and was used to indicate developmental delay. A score of ≤ 70 (two standard deviations below the mean) on the BSITD-III was also assessed to determine how test characteristics differ with a wider range of delay.

4.4 Sample Size

The formula to calculate sample size to determine the precision of the estimate in the diagnostic tests was derived from Jones, Carley and Harrison (2003). Based on limited recruitment, the sample size estimate was estimated with relatively wide parameters, with sensitivity set at 90% and specificity at 80%. Further, the calculation was determined based on a two tail confidence interval at 20% (10% per tail). These

estimates determined that the sample size to evaluate the test characteristics of the tools should include 58 high risk children.

Sensitivity:

$TP+FN=Z^2 [SN (1-SN)] / W^2$

$= 1.96^2 \times [(0.9) (1-0.9)]/0.2^2$

$= 3.8416 \times 0.09/0.04$

$= 8.6436$

Sample Size:

High Risk (SN) = (TP+FN)/P

= 8.64636/0.15 = 58

Specificity:

$TP+FN=Z^2 [SP (1-SP)] / W^2$

$= 1.96^2 \times [(0.8) (1-0.8)]/0.2^2$

$= 3.8416 \times 0.16/0.04$

$= 15.3664$

Sample Size:

High Risk (SP) = (TP+FN)/ (1-P)

= 15.3664/1-0.15 = 19

TP+FN= Sensitivity

SN= Lowest accepted sensitivity: 90%

SP = Lowest accepted specificity: 80%

P: Probability of developmental delay. High risk: 15%

(derived from Vincer et al., 2005)

W: Two-Tailed Confidence interval: 20% (10% each tail)

(Jones et al., 2003).

4.5 Initial Assessment

Prior to the commencement of the study, the administration procedures of the Rourke and NDDS were reviewed by the researcher. Consent forms and study information were mailed to parents approximately one month prior to the scheduled Perinatal Follow-up Program appointment. Parents were given the opportunity to opt out of the study via mail-in card indicating that they were not interested in participating. If parents did not opt out of the study, the researcher called the participant's parents approximately one week prior to the scheduled assessment at the Perinatal Follow-up Program. At this initial phone assessment, participant's parents reviewed the consent form and provided a verbal consent to the administration of the tests. The Rourke Baby Record: Thirty-Six Months was administered to the participant's parent. Minimal probing was provided and parents were encouraged to answer as best describes their child's behaviour. The Rourke Baby Record: Thirty Six Months, developmental section, has nine questions, and parents were instructed to answer yes/no. The recommended method of administration of the Rourke is by a health care professional to the parent. The Rourke and the NDDS were purposely administered prior to the administration of the BSITD-III to avoid influencing the report by parents. If the BSITD-III was performed prior to the administration of the Rourke or NDDS, parents may observe a characteristic that would alter their response. Hence, by administering the Rourke and NDDS first, an unbiased answer was obtained from the parents. This method is consistent with the

underlying mechanisms of a screening test, to predict the need for further investigation with an assessment tool such as the BSITD-III.

4.6 Observed Developmental Assessment

Approximately one week following the Rourke assessment by telephone, participants attended the scheduled Perinatal Follow-up Program appointment. A component of this appointment is the administration of the BSITD-III. Parents were asked to submit the signed consent at this time. By consenting to participation in the study, parents agreed to release the results of the BSITD-III to the researcher.

The NDDS was administered to the parents at the appointment prior to the commencement of the BSITD-III, to ensure that responses would not be biased by the child's performance on the BSITD-III. The NDDS is recommended to be administered via parental scoring of the questionnaire; therefore this method maintains consistency with testing protocols. The NDDS: Thirty-Six Months assessment has eighteen questions to which the parent is to respond yes/no.

Following completion of the NDDS, the BSITD-III was administered by the Perinatal Follow-up Program staff as per protocol within the Perinatal Follow-up Program. Parents were asked to minimize interjections during test administration and to allow the child to make his or her best attempt at the task. They were informed that they could terminate testing at any time as they deemed necessary. Parents were also informed that the test would be scored and results provided following test administration

Test administration was approximately 60-90 minutes in duration, dependant on the child's cooperation and frequency of required breaks. The testing was completed as per the protocol outlined in the BSITD-III administration manual (Bayley, 2006) by the Developmental Associate or Occupational Therapist. A pediatric assessment room at the IWK Health Centre or at the travelling clinic location was used to complete the testing for all participants. The assessment room was located in a quiet location of the hospital with minimal distractions. A child size table and chair was made available, as well as a play area to complete the tests. The room is an appropriate size to allow the parents to be present but not an active participant in the testing. Standardized materials for the administration of the BSITD-III were present for the procedures. The child was permitted to take breaks through the testing and seek comfort from parents as required.

Following the administration of the BSITD-III, parents were provided with the BSITD-III parental questionnaires and asked to complete and mail back to the Follow-up Program. Parents had the opportunity to ask any questions regarding their child's development to the Perinatal Follow-up Team throughout the assessment.

4.7 Resources

A BSITD-III kit is available to the Perinatal Follow-up Program. The Rourke is available on the internet and was downloaded from the website (Rourke, 2009). The NDDS was purchased directly from the NDDS website (NDDS, 2007). Resources were made available to the researcher for training and standardization of methods prior to commencement of the study. The researcher is familiar with administration of pediatric

standardized assessments; hence this served as a benefit to the study as it minimized the required time and resources to train and gain familiarity with the assessments.

4.8 Practical Significance

This pilot study has particular clinical significance in that the Rourke and NDDS are frequently administered to screen for developmental delay in children. However, the ability of these tools to detect developmental delay is largely unknown. Moreover, there is keen interest in identifying a best-practice tool targeting thirty-six month old children. It is anticipated that upon comparison with a gold standard assessment such as the BSITD-III, professionals tracking child development will be provided with information regarding the function of these tools to identify delay in the pediatric population.

4.9 Pilot Studies

This project was conducted as a pilot study, as this method is useful to determine the feasibility of study recruitment and data collection procedures prior to executing a full scale study. Lancaster, Dodd and Williamson (2004) suggested that pilot studies are an important component in preparation of larger scale studies, but are often underutilized and underappreciated, particularly in the published literature. A pilot study has seven key objectives, specifically to assist with sample size calculation, investigation of the integrity of the protocol, evaluation of the data collection forms, collect information regarding recruitment and consent process, determine acceptability of the intervention, review the feasibility of randomization procedures, particularly for pilot studies to randomized controlled trials, and evaluate the appropriateness of the outcome measures (Lancaster et al., 2004).

A preliminary evaluation of testing procedures prior to a larger scale study is useful, particularly with vulnerable populations such as high risk children, as it permits the investigators to carefully review the procedures and appropriateness of evaluation of the tools (Szklo and Nieto, 2007). It is recommended that analysis of pilot studies focus on descriptive analysis to guide the formation of the future project, and significance should not be a deciding factor in the determination to continue to a larger scale study (Lancaster et al., 2004).

4.10 Data Collection

Data collected from the nominal scales (yes/no responses) of the Rourke and NDDS was recorded, with one flag serving as the threshold to warrant further investigation of developmental delay (Rourke, 2007; NDDS, 2007). The data was further assessed at two flags and greater than two flags. The BSITD-III data was collected in a face-to-face direct assessment as required by the standardized test protocol. Results were scored by a computerized program and subdivided into the three categories of motor scale, cognitive scale and language scale. Demographic data were collected from chart review following the Perinatal Follow-up Program appointment. Information related to medical events at birth was also collected, to determine if specific events were indicative of delay. Specific events of interest included neurological event, including seizures, cerebral haemorrhage, hypoxia or ischemic events; cardiovascular events, including cardiac arrest, patent ductus arteriosis or hypotension; respiratory events, including pnuemothorax, pneumonia, idiopathic respiratory distress syndrome, asphyxia, slow to adapt and bronchopulmonary dysplasia. Fetal malnutrition, diagnosed by clinical assessment, typically resulting from insufficient supply of nutrients from the placenta to

the fetus was documented. Information was collected based on indication of the event in the neonatal intensive care unit medical record or discharge summary.

CHAPTER 5: ANALYSIS

The sensitivity, specificity, likelihood ratios, positive predictive value, negative predictive value, false positive rates, false negative rates, likelihood ratios and accuracy were calculated and compared between the Rourke and the NDDS screening tools to the BSITD-III. Children were considered to be at risk of developmental delay with one or more flags on the Rourke. These values were assessed in comparison to the results of the BSITD-III. A similar analysis was completed with the NDDS data in comparison to the BSITD-III. Results were calculated for each participant's Rourke: BSITD-III score, or NDDS: BSITD-III score. Results were not summated over all participants prior to completion of the analysis of the test characteristics.

Further statistical assessment was completed using logistic regression analysis with scoring on the Rourke and the NDDS, in comparison to the outcome BSITD-III to quantify the strength of the association between an abnormal screen and developmental delay. The analysis was then adjusted for gestational age and twin gestation. Pearson correlation estimates were calculated for each of the screening tools' flags relative to the BSITD-III at one and two SD below the mean. Statistical analysis was completed using SAS version 9.2.

Data entry was validated according to double data entry procedures. Upon secondary entry, discrepancies were identified and the chart was reviewed to determine the true result.

CHAPTER 6: RESULTS

6.1 Recruitment

Data was collected between November 12, 2010 and June 15, 2011. Sixty-four children met the inclusion criteria and were scheduled to be seen for assessment with the Perinatal Follow-up Program during the study period. Mail-outs describing the study, consent forms and an opt-out form were sent to these families. Nine families opted out of the study via mailing back the opt-out form. Eleven families were unable to be reached despite multiple calls placed the week prior to the appointment visit. Seven families had disconnected phone numbers with no alternative phone number. Of the 64 children eligible to participate, only 37 parents were reached by telephone. However, 1 did not attend the scheduled appointment, and 5 verbally opted out upon telephone contact. Hence, the recruitment rate of this study was 48.4%. As previously noted, the sample size calculation suggested that 58 high risk children were required for the study to improve the precision of the estimate. The total sample of this study was 31 participants. Refer to Table 1.2 for further information.

6.2 Descriptive Statistics: Participants

Developmental delay, defined as scoring ≤ 85 on any BSITD-III subscale, was observed in 12.9% of children in this sample (95% exact CI: 3.6%-29.8%). Table 1.1 describes demographic and medical characteristics of the study participants. The mean age of participants at the time of administration of the BSITD-III was 37 months, 6 days. Although the evidence is mixed regarding the use of the chronological versus the corrected age at ≤36 months, the Perinatal Follow-up Program abides by procedures outlined in the

BSITD-III manual which suggest the use of the chronological age; hence this is the age to be referred to for the remainder of the document. The average gestational age at delivery of participants was 31 weeks, 6 days. Mean birth weight was 1637.19 grams. 41.94% of participants were twin gestation. 54.84% were male. 80.65% of participants were born prematurely (<37 weeks gestation) and over half (54.84%) were born very prematurely (≤31 weeks gestation). Participants experienced a range of medical conditions at birth, including neurological events (41.94%), such as intraventricular haemorrhage or seizures. 23.33% of participants experienced a cardiac event, such as patent ductus arteriosis or cardiac arrest. 61.29% of participants experienced a respiratory event, most frequently idiopathic respiratory distress syndrome (IRDS) requiring O_2 assistance. Other complications included septicaemia, necrotizing enterocolitis, or inguinal hernia.

6.3 Descriptive Statistics: Screening Tools

Table 2.1 demonstrates the number of flagged areas of concern on the Rourke for the 31 respondents. Of the 31 respondents, 21 (67.7%) indicated that they had no areas of concern on the Rourke. There were 5 parents with one area of concern (16.1%), and 5 parents with two or more areas of concern. The average score on the Rourke was 8.3/9 (range: 4-9), indicating that few parents flagged areas of concern. Five parents answered positive for the last question, "No parental/caregiver concerns". With this question removed the average score on the Rourke was 8.4/9 (range: 5-9). With the criteria set at one question indicating an area concern as being indicative of a positive screen, and recommending referral for further assessment, ten participants screened positive on the Rourke, reduced to nine participants once the final question was removed from the analysis. Both scores will be assessed in relation to the BSITD-III herein as this question is not directly related to a

child's developmental performance, although it appears that despite its subjectivity it had minimal impact on the results. The other question that was frequently flagged on the Rourke was "Does your child use sentences with 5 or more words?" flagged in 19.4% of participants. Refer to Table 2.1 for further information.

Table 2.2 demonstrates the number of flagged areas of concern on the NDDS for the 31 respondents. Of the 31 respondents, 22 (70.0%) indicated that they had no areas of concern on the NDDS. There were 5 parents with one area of concern (16.1%) and 4 parents (12.9%) with two or more areas of concern. The average score on the NDDS was 17.1/18 (range 9-18), again indicating that the majority of parents did not flag areas of concern. With the criteria set at one question indicating an area of concern as being indicative of a positive screen, and recommending referral for further assessment, nine participants screened positive on the NDDS. On the NDDS, the two questions most frequently flagged included, "Does your child speak clearly enough to be understood most of the time by family?", flagged by 12.90% of participants and "Does your child dress and undress with help?", again flagged by 12.90% of participants. Refer to Table 2.2 for further information.

6.4 Descriptive Statistics: BSITD-III

The findings for the BSITD-III are found in Table 3.0 and 4.0. Of the 31 respondents, 27 (87.1%) scored within one SD from the mean on all composite scores of the BSITD-III. There were 2 (6.5%) participants who scored below one SD of the mean and 2 that scored below two SD of the mean. Table 3.0 illustrates the composite scores for the motor, cognitive and language scales. Table 4.0 illustrates the findings for analysis for each of the subscales: fine motor, gross motor, cognitive, expressive language and receptive

language. It indicates the mean scaled score of the participants and the mean age of developmental performance of the participants. Given the considerably low response rate for parental questionnaires (1/31), this response was omitted from the analysis.

6.5 Test Characteristics of Developmental Screening Tools

Findings pertaining to the test characteristics of the tools can be found in Tables 5.1 and 5.2. Both the Rourke and NDDS were assessed using one or two flags as the cut-off criteria. As previously noted, a flag is indicative of an area of concern reported by the parent, and is used to indicate referral for further assessment. Further, the Rourke score with the omission of the open-ended question pertaining to parental concerns was also assessed using the flag criteria.

Table 5.1.i. demonstrates the results of performance on the Rourke with one flagged area of concern in comparison to the BSITD-III when delay was marked by at least one scale (motor, language or cognitive) one standard deviation below the mean (i.e., a score of $\leq$85). Results indicate sensitivity was 75%, specificity of 74%, positive predictive value of 30%, negative predictive value of 95%, false positive rate of 26%, false negative rate of 25%, and likelihood ratio of 2.90. Overall accuracy was found to be 74%.

Table 5.1..ii demonstrates the comparison of the Rourke with omission of the last question, an open-ended question regarding parental concerns, at one flagged area of concern compared to the BSITD-III at one SD below the mean. Sensitivity was unchanged at 75%, specificity improved slightly to 78%, positive predictive value to 33%, negative predictive value of 95%. Rate of false positives dropped slightly to 22%, there was no

change in false negative rate of 25% and the likelihood ratio improved to 3.41. Overall accuracy was 77%.

When set to two flagged areas of concern on the screening test (Table 5.1.iii), the results of the Rourke in comparison to the BSITD-III (one SD below mean) had a sensitivity of 75%, specificity of 93%, positive predictive value of 60%, negative predictive value of 96%, false positive rate of 7%, false negative rate of 25%, and a likelihood ratio raised to 10.71. Overall accuracy was 90%. Analysis of the Rourke with omission of the last question at two flagged areas of concern in comparison to the BSITD-III at one SD below the mean (Table 5.1.iv) demonstrated identical results.

Results of the comparison of the NDDS at one flagged area of concern in comparison to the BSITD at one SD below the mean , as indicated in Table 5.1.v., found sensitivity was 75%, specificity of 78%, positive predictive value of 33%, negative predictive value of 95%, false positive rate at 22%, false negative rate of 25%, and likelihood ratio of 3.41. Overall accuracy was 77%. Upon comparison of the NDDS with two flagged areas of concern, to the BSITD-III (1 SD below the mean) demonstrated in Table 5.1.vi, sensitivity was 75%, specificity of 96%, positive predictive value of 75%, negative predictive value of 96%, false positive rate at 4%, false negative rate of 25%, and likelihood ratio of 20.27. Overall accuracy was found to be 94%.

Table 5.2.i demonstrated the results of the Rourke at one flagged area of concern compared to performance on the BSITD-III at two standard deviations below the mean (≤70). Results found that sensitivity was 100%, specificity of 72%, positive predictive value of 20%, negative predictive value of 100%, false positive rate at 28%, false negative rate of

0% and likelihood ratio of 3.62. Overall accuracy was 74%. Upon omission of the final question on the Rourke pertaining to parental concern, at one flagged area of concern compared to the BSITD-III at two SD below the mean (Table 5.2.ii), sensitivity remained at 100%, specificity improved to 76%, positive predictive value improved slightly to 22%, negative predictive value was 100%, false positive rate decreased to 24%, false negative rate was unchanged at 0% and the likelihood ratio improved to 4.15. Overall accuracy improved to 77%.

When the Rourke were assessed using the two flagged areas of concern compared to the BSITD-III score at two standard deviations below the mean, as demonstrated in Table 5.2.iii, it demonstrated sensitivity of 100%, specificity of 90%, positive predictive value of 40%, negative predictive value of 100%, false positive rate of 10%, false negative rate of 0% and likelihood ratio of 10.34. Overall accuracy was 90%. Analysis of the Rourke with omission of the last question at two flagged areas of concern in comparison to the BSITD-III at two SD below the mean (Table 5.2.iv) demonstrated identical results.

For the NDDS implementing the one flagged area of concern cut point, in comparison to the BSITD-III score with two SD below the mean, as demonstrated in Table 5.2.v., sensitivity was 100%, specificity of 76%, positive predictive value of 22%, negative predictive value of 100%, false positive rate at 24%, false negative rate of 0% and likelihood ratio of 4.15. Overall accuracy was 77%.

For the NDDS implementing the two flagged areas of concern cut point in comparison to the BSITD-III score with two SD below the mean (Table 5.2.vi), sensitivity was 100%, specificity of 93%, positive predictive value of 50%, negative predictive value of

100%, false positive rate at 7%, false negative rate of 0% and likelihood ratio of 14.50. Overall accuracy was 94%. The small cell sizes, particularly the zero cell value in the BSITD-III at two SD below the mean, reduced the capacity for further assessment; hence subsequent analysis was completed in relation to the BSITD-III at one SD below the mean.

6.6 Logistic Regression Analysis

Logistic regression analysis was completed to determine if the score on the screening tool was significantly related to the BSITD-III at one SD below the mean. Models were conducted for one and two flag cut points for the screening tools. No flags, hence a negative screen for developmental delay, was the referent category. Each analysis was adjusted for twin gestation and gestational age. However, due to the small cell sizes, this adjustment did not produce meaningful results and findings remain inconclusive. These values are demonstrated in Table 6.0 for further information.

Upon analysis of the Rourke at one flag cut points, unadjusted analyses suggests that there is no significant association of the Rourke to a positive assessment indicative of delay on the BSITD-III, likely due to low power (1 Flag: OR: 8.57, 95% CI: 0.76-96.52) as well as the NDDS (1 Flag: OR: 10.50, 95% CI: 0.92-120.26). It was not possible to model the adjustment at two flag cut points due to low cell size.

A Pearson correlation matrix was completed for each of the flag cut-points noted to the BSITD-III. Results are shown in Table 7.0. Results suggest a relationship between flags on the screening test and to whether the child scored one or two SD below the mean on the BSITD-III. Results suggest a positive association between the flags on both the Rourke and the NDDS relative to the BSITD-III, suggesting that as number of flags increased,

indicating greater number of parental concerns pertaining to development, the child was more likely to have a BSITD-III composite score below the mean. There was a slightly stronger correlation of the NDDS to the BSITD-III score in comparison to the Rourke at both one and two SD below the mean [e.g. r=0.71 (Rourke), r=0.85 (NDDS) at more than two areas of concern on the screen tool relative to the BSITD-III].

CHAPTER 7: DISCUSSION

Developmental delay was evident in 12.9% (95% exact CI: 3.6%-29.8%) of this study's population. Although lower than the reported rates of developmental delay, which suggest that children born prematurely have rates as high as 23%, it should be noted that the criteria for admission to this study was a higher birth weight (≤1500 grams compared to 800 grams) and an older gestational age to classify prematurity (≤31 weeks compared to <26 weeks gestation) in comparison to a previous study (Lorenz et al, 1998). Conversely, Boyle et al (1994) suggested that up to 17% of children experience some form of developmental delay, including learning disability, by 18 years old. However, even if the overall rate was lower than previously reported rates, the confidence interval of this study's finding encompasses the range of previous studies, thereby demonstrating some level of consistency.

Participants in this study were approximately three years old and may not yet demonstrate higher order function which would present as developmental delay as they age and prepare for school. Further, these children have been monitored regularly by the Perinatal Follow-up Program since birth, and therapeutic interventions for developmental delay may have ameliorated potential expected delays.

In terms of screening tool properties, analysis suggests that the Rourke and the NDDS tests had excellent sensitivity and negative predictive values. Investigation of the ability of these tests to detect developmental delay had not previously been completed; therefore these findings suggest that the tools may be appropriately screening for developmental delay with very few false negatives. The majority of children who tested positive for developmental

delay on the BSITD-III also received a positive result on one of the screens. This suggests that the screening tools are indeed sensitive enough to detect delay when it is present.

However, the false positive rate varied between the tools, and some children who screened positive were found to test negative for developmental delay. This suggests that the tests may be increasing referral rate for further assessment when it is not warranted. However, given that specificity rate was also high, it is suggested that over estimating the potential children with developmental delay is a wiser move from a health care perspective given that it will ultimately results in slightly more assessments at the cost of ensuring that all children are appropriately tested for developmental delay.

Interestingly, of the three questions flagged in the Rourke and NDDS of potential areas of concern, two pertained to the child's language production, even though receptive and expressive communication scored slightly higher compared to the other two subscales on the BSITD-III. One possible explanation may be related to clarity of speech production, in comparison to the content of the speech, the component assessed in terms of receptive and expressive form by the health professionals. Secondly, there were five participants who had a flagged area of concern on either the Rourke or the NDDS, but not both screening tools. Upon further analysis of these questions, of the three participants who flagged positive on the Rourke, two were pertaining to the child's ability to speak in sentences in five or more words (the third participant indicated a general area of parental concern on question nine). A similar question regarding language production is located on the NDDS, but asks if the child can speak 2-5 word sentences. This discrepancy in response highlights the acuteness of the parents' response, and the importance of appropriately defining developmental milestones in the screening assessments. In terms of flagged areas of concern noted on the

NDDS but not the Rourke, two participants indicated concern regarding gross motor skills, specifically tossing a ball and getting dressed with assistance. These two skills do not have a similar counterpart on the Rourke, further suggesting how performance on a developmental screening test can differ depending on how a domain of child development is assessed.

The findings of the Pearson correlation matrix suggest a relationship between the screening test flagged score and if the child scored one or two SD below the mean on the BSITD-III. This finding was observed across all flag cut points. Interpretation of these findings is limited given that the tools were assessed at cut points and not as a continuous variable. However, it demonstrates a trend that the screening tool results appear to be related to a negative test on the BSITD-III.

As noted, the lack of significance in the logistic regression analysis of the screening tests is likely attributed to the small sample size. The odds ratios results increased from the two flag cut point relative to the one flag cut point for each of the screening tests in the unadjusted analysis. This is expected, as an individual is more likely to have developmental delay if they demonstrate more flagged areas of concern. Significance at the two or more flags cut point was demonstrated for each of the screening tools. However, given the wide confidence intervals it is likely that this is the result of a type 1 error. Further, the odds ratios did not appear to be appropriate estimates of the relative risk due to the relatively high proportion of children with developmental delay. If the occurrence of developmental delay was rare, the relative risk may be an appropriate estimate of the odds ratio. As noted, the small cell sizes contributed to the poor fit of the model once adjusted for potential confounders, and therefore did not contribute meaningful results.

The Rourke's closed-ended question pertaining to general parental concern appeared to have a minor effect on the results, but only at the one flag level at one SD below the mean on the BSITD-III. Review of the results suggests that the majority of the parents who are reporting concerns regarding their child's development are also indicating areas of concern on other areas of the Rourke. However, only approximately half of the participants with noted parental concern regarding development actually tested positive for delay on the BSITD-III. Once this question was removed from the analysis, the false positive rates were reduced slightly from 26% to 22%. As this is not a considerable difference, it suggests that this question is not significantly altering the test properties of the Rourke. However, further investigation should continue to assess the inclusion of this question and its contribution to the findings.

In terms of the false positive rate, which is reflective of whether children who do not have developmental delay are screened positive, the majority of the assessments had a rate ranging from 22-28% when assessed at the one flag level. However, when assessed at two flags for each of the Rourke, Rourke minus question nine and the NDDS, in relation to both one and two SD below the BSITD-III mean, there was a drop in the false positive rate to 4-10%, suggesting that the increase in flagged concerns for screening decreased the risk of a child's being further assessed for delay when it is not warranted. False negative rate, indicating that children who have developmental delay are screening negative dropped from 25% to 0% when the BSITD-III cut point changed from one to two standard deviations below the mean for all screening tools. Although this change was indicated based on the results of one participant, it reiterates the change in test characteristics based on screening test cut points.

As previously noted, the Rourke is designed to serve as a surveillance measure, not a screening tool. Hence, its ability to screen for developmental delay is dependent on assessment over time ("Developmental Surveillance", 2001). Further, it includes information from other measurements, such as physical examination as well as parental report. In this pilot study, the Rourke was used as a screening tool, in that it assessed the child's development at a specific time point. Future studies of the Rourke's ability to detect delay should include information collected from the Rourke as a surveillance measure, to more accurately depict the appropriate administration methods of the measure.

The importance of the cut point value for developmental delay is an important consideration. Dahinten &Ford (2004) assessed performance of the NDDS at one and two flags, relative to 1, 1.5 and 2 SD below the mean on the BSID-II. Results varied in comparison to those found in this study. For example, for the NDDS one flag cut point, at one SD below the mean BSID-II, they found sensitivity at 56.3% (compared to this study with sensitivity at 75%) and specificity at 70.6% (compared to 78%). The results are similar when expanded to two SD below the BSID-II mean, although the sensitivities were identical at 100%, they identified specificity of 68.7% compared to 76% in this study. Of particular note, the false positive, hence potential over-referral rate also differed between the two studies. They identified the rate at 25.4% for NDDS one flag, one SD below the BSID-II mean (compared to 22% in this study) and 30.5% at two SD below the BSID-II mean, compared to 24% in this study.

The discrepancy between the results, with Dahinten &Ford (2004) obtaining lower values in the test characteristics compared to this study, continued with NDDS assessed at the two flag cut point. At one SD below the mean on the BSID-II, they found sensitivity of

43.8% (compared to 75%) and specificity at 86.3%, lower than from this study's finding of 96%. When expanded to two SD below the mean on the BSID-II, again the sensitivity values were identical across the two studies, and specificity at 84.3% (compared to 93%). They suggested that the NDDS was less precise of a measurement than the findings of this study suggest. A possible explanation may be due to differences in methodology. Dahinten & Ford (2004) included a younger population, ranging in age from 4-24 months, when it may be more difficult to detect delay as the child may not have reached an older age to demonstrate or solidify development of a particular milestone, which may be more apparent at 36 months. Further, the range of ages could contribute to the decreased precision of the measurement. In addition, Dahinten &Ford (2004) assessed children from the general population, which may contribute to decreased ability to detect delay than this study's high risk population with greater susceptibility to delay. Further, the discrepancies were greater at the one SD below the BSID-II mean than at two SD below the mean, suggesting that the NDDS may have improved test characteristics to assess more profound delay in a general population than when it has to detect subtler delay at only one SD below the mean. Finally, the Dahinten &Ford sample included 118 participants, compared to 31 in this study, which would likely contribute to discrepancies, as noted above in terms of the difference in sensitivity values. A larger sample inherently demonstrates improved representativeness of a population.

Despite the attempt to control for various levels of bias, several potential sources remain within the study. There was a risk of an effect of testing order given that the sequence of administration of the Rourke and the NDDS wasn't randomized. However, there was likely limited impact of testing effects as the two series of testing are measured via different

methods (parental observation versus behavioural assessment) by two different individuals. Further, all tests are standardized, limiting subjective interpretation. Test-retest concerns are minimized as the BSITD-III, NDDS and Rourke are administered once. The threats of maturation effects are present but limited between the first and second assessment due to the limited time frame for assessment (within one week). Follow-up at thirty-six months is a standard procedure within the IWK Health Centre for high risk children. Therefore, this decreases the representativeness of the sample to any child at high risk of delay who may not be receiving active treatment, and may demonstrate worse characteristics than those without follow-up. Ethical considerations do not permit a high risk child to be allocated to a group that does not receive follow-up. However, given that these children are receiving regular follow-up by a neonatologist and potentially other members of the health care team (e.g. occupational therapist, physiotherapist, speech language pathologist), this group may demonstrate an improvement in skills due to the advanced implementation of early intervention services, and therefore may not demonstrate skills as impaired as expected in a high risk population. Also, parents of children at high risk of developmental delay may not be representative of a general parental population, thereby impacting their report of their child's development. Parents of these children may be hyper-vigilant in observation of their child's behaviour and development given that their child has had regular monitoring for delay since birth. Therefore, they may over-report delay when the characteristic isn't truly due to developmental delay. However, these parents may provide self report with increased accuracy given their history of reporting their child's development to health professionals, hence may provide a more accurate self report compared to the general population. A qualitative analysis of a parental response may provide insight to this issue.

A further threat to the representativeness of this sample to the population of all children at high risk of developmental delay is adherence with participation with the follow-up program. Research suggests that high risk children who do not regularly attend appointments, defined as two or missed appointments over a two year period, due to factors such as social mobility, had a 5% increase in rate of developmental delay than children who regularly attended appointments. Further, children who did not attend appointments and were removed from parental care were at eight times greater risk of delay. At 36 months, the families who continue to participate in the program have found means to accommodate appointment visits and necessary travel, which may suggest that individuals with decreased access to resources may have lower adherence to appointment schedules. Hence, the participants may have demonstrated lower levels of disability than a child at high risk of developmental delay due to their ability to continue to attend appointments and receive treatment (Tin et al., 1998).

Of the four children with scores indicative of delay on the BSITD-III, all were boys. This is consistent with the literature which suggests that boys tend to have a higher rate of developmental delay than girls (deMoura, Costa, Santos, Barros, Matijasevich, et al., 2010; Lai, Tseng, & Guo, 2011). Upon analysis of the BSITD-III, as reflected in Tables 3.0 and 4.0, few children demonstrated scores which would be indicative of developmental delay. Anderson et al (2010) suggested that the BSITD-III may underestimate developmental delay. They suggest that the reference values used to indicate normal development are not representative of true performance in a general population. When they compared two samples of children, one cohort at elevated risk of delay and one control, the reference values provided by the BSITD-III suggested no true

difference in performance, but actual performance between the two groups were suggestive of delay. In consideration of this finding, participants in this study may have had delay which, theoretically, may have been identified on the screening tool but not on the BSITD-III. This would have important clinical implications, and should be considered a possible limitation of the study as well as the assessment process for pediatric development (Anderson et al., 2010).

A strength of this study is that it provided insight on the performance capabilities of assessment tools frequently administered in clinical practice to screen for developmental delay in children. It provides support for further analysis within a larger, sufficiently powered study. The findings suggest that results of a larger study could potentially alter professional practice. A limitation of this study is that all children assessed were, by definition of being enrolled in the Perinatal Follow-up Program, at high risk of developmental delay. Therefore, all children were receiving regular medical services, which limit the generalizability of the findings. Secondly, the low response rate (48 %) reduced the precision of the estimates of the screening characteristics. Further, it may have reduced the representativeness of the sample.

CHAPTER 8: CONCLUSION

Upon initiation of a full scale study utilizing the tested study procedures, the following adjustments to the protocol are proposed. First, it is suggested that children at low risk of developmental delay be included within the sample to determine if the tools can appropriately screen for developmental delay within a population with lower probability of occurrence. Secondly, it is recommended that administration of the screening tool be completed during the follow-up program appointment, prior to completion of the BSITD-III. An inability to reach parents by telephone to complete the Rourke prior to the follow-up appointment impacted the response rate, and may be overcome by this adjustment. Thirdly, the sample size calculation for this assessment was determined to provide a level of precision of +/- 10% for the estimated level of sensitivity. Future studies should decrease this value to +/- 5% to improve the level of precision of the analysis, thereby providing stronger evidence regarding the tool's ability to screen for developmental delay. Finally, it is recommended that the screening tools be administered in a randomized order to prevent the opportunity for a parent to be influenced by the response on the previous tool.

As previously noted, in Ontario it is recommended that the NDDS be administered in conjunction to the Rourke to provide a comprehensive screening result (Rourke, 2009). It is speculated that given the similarities between the tools, administering both tools at the same appointment would not provide appropriate results, as parents may defer to their answer to a similar question asked on the earlier screening test, versus careful consideration of how the questions may differ. However, administration of the two tests at proximal time points, to ensure that the same developmental stage is assessed, but not at the same visit, may be beneficial. Within this study it was noted that five participants were

flagged as an area of concern on either the Rourke or the NDDS, but not both. Upon further analysis, it appeared that the questions that differed were assessing different domains, for example, can the child speak 5 or more words, as assessed on the Rourke, versus whether the child can speak 2-5 words, as assessed on the NDDS. Together, the two screening tests provide a more comprehensive analysis of the child's verbal communication skills. However, if both tools were to be administered to a child, each test score should be reviewed independently and careful consideration of how the domain is related to the child's development, as evaluation of the overall assessment score alone may suggest elevated concern for delay when in fact the child is scoring as potentially impaired on two similar, or even identical, domains.

The low sample size of this study, coupled with the inherent limitations of a pilot study (Lancaster et al., 2004), limits our ability to draw conclusions from the findings. However, it appears that the modified Rourke and the NDDS provide reasonably good screening characteristics for developmental delay upon comparison to the BSITD-III. Further, screening properties of the tools are improved when the criteria for developmental delay extended to two flagged areas of concern on the screening tests. A future larger, sufficiently powered study should examine the effect of the screening tools on a more diverse population by confirming the results in comparison to children at low risk of developmental delay. The scope of this research topic has the potential to alter clinical practice as it suggests that the Rourke and the NDDS demonstrate similar abilities to screen for developmental delay. Therefore the results of this study should be further investigated with the appropriate adjustments to further investigate the properties of the tools.

REFERENCE LIST

Anderson, P.J. De Luca, C.R., Hutchinson, E., Roberts, G., Doyle, L.W. and the Victorian Infant Collaborative Group (2010). Underestimation of developmental delay by the new Bayley-III scale. Archives of Pediatric and Adolescent Medicine, 164 (4), 352-356.

Aylward, G. (2009). Developmental screening and assessment: what are we thinking? *Journal of Developmental and Behavioural Paediatrics*, 30 (2), 169-173.

Bayley, N. (2006). Bayley Scales of Infant and Toddler Development, Administration Manual (3rd edition). San Antonio, Tx. Harcourt Assessment Inc.

Bayley, N. (2006). Bayley Scales of Infant and Toddler Development, Technical Manual (3rd edition). San Antonio, Tx. Harcourt Assessment Inc.

Bracken, B.A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. Journal of Psychoeducational Assessment, 5(4):313.

Boyle CA, Decoufle P, Yeargin-Allsopp M. (2004). Prevalence and health impact developmental disabilities in US children. *Pediatrics* 93(3):399-403.

Canadian Paediatric Society (2008). Publications and Resources: *The Rourke Baby Record*. Retrieved November 18, 2009 from http://www.cps.ca/english/statements/cp/rourke/rourkebabyrecord.htm

Catto-Smith, A. G., Yu, V. Y., Bajuk, B., Orgill, A. A., & Astbury, J. (1985). Effect of neonatal periventricular haemorrhage on neurodevelopmental outcome. *Archives of Disease in Childhood, 60*(1), 8-11.

Clarke, M., Hague, D., Mortimer, B., & Thompson, S. (2004). Should every three-year-old receive a routine review? *Community Practitioner, 77*(3), 101-104. Retrieved from http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h& AN=2004077078&site=ehost-live.

Connolly, B. H., Dalton, L., Smith, J. B., Lamberth, N. G., McCay, B., & Murphy, W. (2006). Concurrent validity of the bayley scales of infant development II (BSITD-II) motor scale and the peabody developmental motor scale II (PDMS-2) in 12-month-old infants. *Pediatric Physical Therapy : The Official Publication of the Section on Pediatrics of the American Physical Therapy Association, 18*(3), 190-196.

Cossette L and Duclos E. A profile of disability in Canada, 2001. Retrieved on April 14, 2010 from http://www.statcan.gc.ca/pub/89-577-x/pdf/4228016-eng.pdf.

Dahinten, V.S. and Ford. L. (2004). Validation of the nipissing district developmental screen for use with infants and toddlers - working paper. *Consortium for health, intervention, learning and development*, University of British Columbia.

de Moura, D. R., Costa, J. C., Santos, I. S., Barros, A. J., Matijasevich, A., Halpern, R., et al. (2010). Risk factors for suspected developmental delay at age 2 years in a Brazilian birth cohort. *Paediatric and Perinatal Epidemiology, 24*(3), 211-221.

Developmental surveillance and screening of infants and young children (2001). *Pediatrics, 108*(1), 192-196.

Dezoete, J. A., MacArthur, B. A., & Tuck, B. (2003). Prediction of bayley and stanford-binet scores with a group of very low birth weight children. *Child: Care, Health and Development, 29*(5), 367-372.

Dinkevich, E., Hupert, J., & Moyer, V. A. (2001). Evidence based well child care. *BMJ (Clinical Research Ed.), 323*(7317), 846-849.

Doig, K. B., Macias, M. M., Saylor, C. F., Craver, J. R., & Ingram, P. E. (1999). The child development inventory: A developmental outcome measure for follow-up of the high-risk infant. *The Journal of Pediatrics, 135*(3), 358-362.

Dupont, W.D. and Plummer W.D. (2009). PS Power and Sample Size Calculation, version 3.0.12. Vanderbilt University.

Evensen K.A., Skranes J., Brubakk A.M., Vik T.(2009).Predictive value of early motor evaluation in preterm very low birth weight and term small for gestational age children. Early Human Development, 85 (8), 511-8.

Field, T., Hallock, N., Ting, G., Dempsey, J., Dabiri, C., & Shuman, H. H. (1978). A first-year follow-up of high-risk infants: Formulating a cumulative risk index. *Child Development, 49*(1), 119-131.

Foster-Cohen, S. H., Friesen, M. D., Champion, P. R., & Woodward, L. J. (2010). High prevalence/low severity language delay in preschool children born very preterm. *Journal of Developmental and Behavioral Pediatrics: JDBP, 31*(8), 658-667.

Fox, N. A., & Porges, S. W. (1985). The relation between neonatal heart period patterns and developmental outcome. *Child Development, 56*(1), 28-37.

Goyen, T. A., & Lui, K. (2002). Longitudinal motor development of "apparently normal" high-risk infants at 18 months, 3 and 5 years. *Early Human Development, 70*(1-2), 103-115.

Hanley, J.A., Negassa, A., Edwardes, M.D. & Forrester, J.E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. American Journal of Epidemiology, 157, 4, 364-375.

Harris, S. R., & Langkamp, D. L. (1994). Predictive value of the bayley mental scale in the early detection of cognitive delays in high-risk infants. *Journal of Perinatology: Official Journal of the California Perinatal Association, 14*(4), 275-279.

Harris, S. R., Megens, A. M., Backman, C. L., & Hayes, V. E. (2005). Stability of the bayley II scales of infant development in a sample of low-risk and high-risk infants. *Developmental Medicine and Child Neurology, 47*(12), 820-823.

Hollingshead, AB. (1975). Holllingshead Index: The Four Factor Index of Social Status. Retrieved May 7, 2010 from http://www.yale.edu/sociology/faculty/docs /hollingshead_socStat4factor.pdf.

Identifying Infants and Young Children with Developmental Disorders in the Medical Home: An Algorithm for Developmental Surveillance and Screening. (2006). *Pediatrics*, 118(1), 405-420.

Isotani, S. M., Azevedo, M. F., Chiari, B. M., & Perissinoto, J. (2009). Expressive language of two year-old pre-term and full-term children. *Pro-Fono: Revista De Atualizacao Cientifica, 21*(2), 155-159.

Janssen, A. J., Nijhuis-van der Sanden, M. W., Akkermans, R. P., Oostendorp, R. A., & Kollee, L. A. (2008). Influence of behaviour and risk factors on motor performance in preterm infants at age 2 to 3 years. *Developmental Medicine and Child Neurology, 50*(12), 926-931.

Janssen,A.J., Nijhuis-van der Sanden M.W., Akkermans R.P., Tissingh J., Oostendorp R.A., Kollée .LA. (2009). A model to predict motor performance in preterm infants at 5 years.*Early Human Development*, 85(9), 599-604.

Jones, S.R., Carley, S. and Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal,* 20, 453-458.

Kono, Y., Mishina, J., Sato, N., Watanabe, T., & Honma, Y. (2008). Developmental characteristics of very low-birth weight infants at 18 months' corrected age according to birth weight. *Pediatrics International: Official Journal of the Japan Pediatric Society, 50*(1), 23-28.

Koseck, K. (1999). Review and evaluation of psychometric properties of the revised bayley scales of infant development. *Pediatric Physical Therapy, 11*(4), 198-204. Retrieved from http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=2000012629&site=ehost-live.

Koseck, K. And Harris, S.R. (2004) Changes in performance over time on the bayley scales of infant development II when administered to infants at high risk of developmental disabilities. *Pediatric Physical Therapy, 16*, 199-205.

Lai, D. C., Tseng, Y. C., & Guo, H. R. (2011). Gender and geographic differences in developmental delays among young children: Analysis of the data from the national registry in Taiwan. *Research in Developmental Disabilities, 32*(1), 63-69.

Lancaster, G.A., Dodd, S., and Williamson, P.R. (2004). Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice,* 10, 2, 307-312.

Leduc, D., Rourke, L., Rourke, J., & Constantin, E. (2006). Health supervision from zero to five years using the 2006 Rourke. *Paediatrics & Child Health, 11*(8), 487-488.

Leib, S. A., Benfield, D. G., & Guidubaldi, J. (1980). Effects of early intervention and stimulation on the preterm infant. *Pediatrics, 66*(1), 83-90.

Lemoine, M. H., Lemoine, C., & Cyr, C. (2006). Paediatric primary care quality and accessibility: Parents' perspective. *Paediatrics & Child Health, 11*(1), 25-28.

Limbos, M. M., & Joyce, D. P. (1999). Developmental screening using the Rourke. *Canadian Family Physician, 45*, 869-870.

Limbos, M., Joyce, D. And Roberts, G. (2010). Nipissing District Developmental Screen: patterns of use by physicians in Ontario. Canadian Family Physician, 56(2), 66-72.

Lorenz, J. M., Wooliever, D. E., Jetton, J. R., & Paneth, N. (1998). A quantitative review of mortality and developmental disability in extremely premature newborns. *Archives of Pediatrics & Adolescent Medicine, 152*(5), 425-435.

Macias, M. M., Saylor, C. F., Greer, M. K., Charles, J. M., Bell, N., & Katikaneni, L. D. (1998). Infant screening: The usefulness of the bayley infant neurodevelopmental screener and the clinical adaptive Test/Clinical linguistic auditory milestone scale. *Journal of Developmental and Behavioral Pediatrics, 19*(3), 155-161.

Nipissing District Developmental Screen (2000). Nipissing District Developmental Screen Intellectual Property Association. North Bay: NDDS.

Nipissing District Developmental Screen (2007). Retrieved November 22, 2009 from http://www.ndds.ca/.

Nipissing District Developmental Screen (2007). *History,* Retrieved November 22, 2009 from http://www.ndds.ca/history.

Ohgi, S., Fukuda, M., Akiyama, T., & Gima, H. (2004). Effect of an early intervention programme on low birth weight infants with cerebral injuries. *Journal of Paediatrics and Child Health, 40*(12), 689-695.

Panagiotou, L., Rourke, L. L., Rourke, J. T., Wakefield, J. G., & Winfield, D. (1998). Evidence-based well-baby care. part 1: Overview of the next generation of the Rourke. *Canadian Family Physician, 44*, 558-567.

Panagiotou, L., Rourke, L. L., Rourke, J. T., Wakefield, J. G., & Winfield, D. (1998). Evidence-based well-baby care. part 2: Education and advice section of the next generation of the Rourke. *Canadian Family Physician Medecin De Famille Canadien, 44*, 568-572.

Pearson Assessment and Administration (2009) Bayley *Scales of Infant and Toddler Development, Third Edition (Bayley-III).* Retrieved November 18, 2009 from http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X&Mode=resource.

Personal Communication (2009). Dr. Sarah Shea, Pediatrician, IWK Health Centre.

Powell, J. and Smith, C.A. (1994). Developmental milestones: A guide for parents. National Network for Child Care - NNCC. The 2nd year. In Manhattan, KS: Kansas State University Cooperative Extension Service. Retrieved December 12, 2009 from http://www.nncc.org/Child.Dev/mile2.html#anchor789390.

Rourke (2009) *Current Evidence.* Retrieved April 9, 2010 from http://www.rourkebabyrecord.ca/evidence.html.

Rourke (2009) *Rourke: Evidence-Based Infant/Child Health Maintenance Guide.* Retrieved April 10, 2010 from http://www.rourkebabyrecord.ca.

Rourke (2009). Changes in the 2009 Rourke Baby Record. Retrieved August 2, 2011 from http://www.rourkebabyrecord.ca/changes.html.

Rourke, L., Godwin, M., Rourke, J., Pearce, S., & Bean, J. (2009). The Rourke infant/child maintenance guide: Do doctors use it, do they find it useful, and does using it improve their well-baby visit records? *BMC Family Practice, 10* (28), 10-28

Scheiner, A. P., & Sexton, M. E. (1991). Prediction of developmental outcome using a perinatal risk inventory. *Pediatrics, 88*(6), 1135-1143.

Singer, L. T., Arendt, R., Minnes, S., Farkas, K., Salvator, A., Kirchner, H. L., et al. (2002). Cognitive and motor outcomes of cocaine-exposed infants. *JAMA: The Journal of the American Medical Association, 287*(15), 1952-1960.

Snow, A.V. and Lecavalier, L. (2008). Sensitivity and specificity of the Modified Checklist for Autism in Toddlers and the Social Communication Questionnaire in preschoolers suspected of having pervasive developmental disorders. Autism. 12(6):627-44.

Szklo, M. And Nieto, F.J. (2007). Epidemiology: Beyond the Basics, 2nd ed. Jones and Bartlett Publishers, Ontario, Canada.

Taylor, J. A., Davis, R. L., & Kemper, K. J. (1997). A randomized controlled trial of group versus individual well child care for high-risk children: Maternal-child interaction and developmental outcomes. *Pediatrics, 99*(6), E9.

Tin W, Fritz S, Wariyar U, Hey E (1998). Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty. Archives of disease in childhood. Fetal and neonatal edition, British Pediatric Association, 79:F83–F87

Vincer, M.J., Cake, H., Graven, M., Dodds, L., McHugh, S. and Fraboni, T. (2005). A population-based study to determine the performance of the cognitive adaptive test/clinical linguistic and auditory milestone scale to predict the mental development index at 18 months on the bayley scales of infant development-II in very preterm infants, *Pediatrics, 116* (6), 864-867.

Woodward, L. J., Moor, S., Hood, K. M., Champion, P. R., Foster-Cohen, S., Inder, T. E., et al. (2009). Very preterm children show impairments across multiple neurodevelopmental domains by age 4 years. *Archives of Disease in Childhood.Fetal and Neonatal Edition, 94*(5), 339-44.

APPENDICES AND TABLES

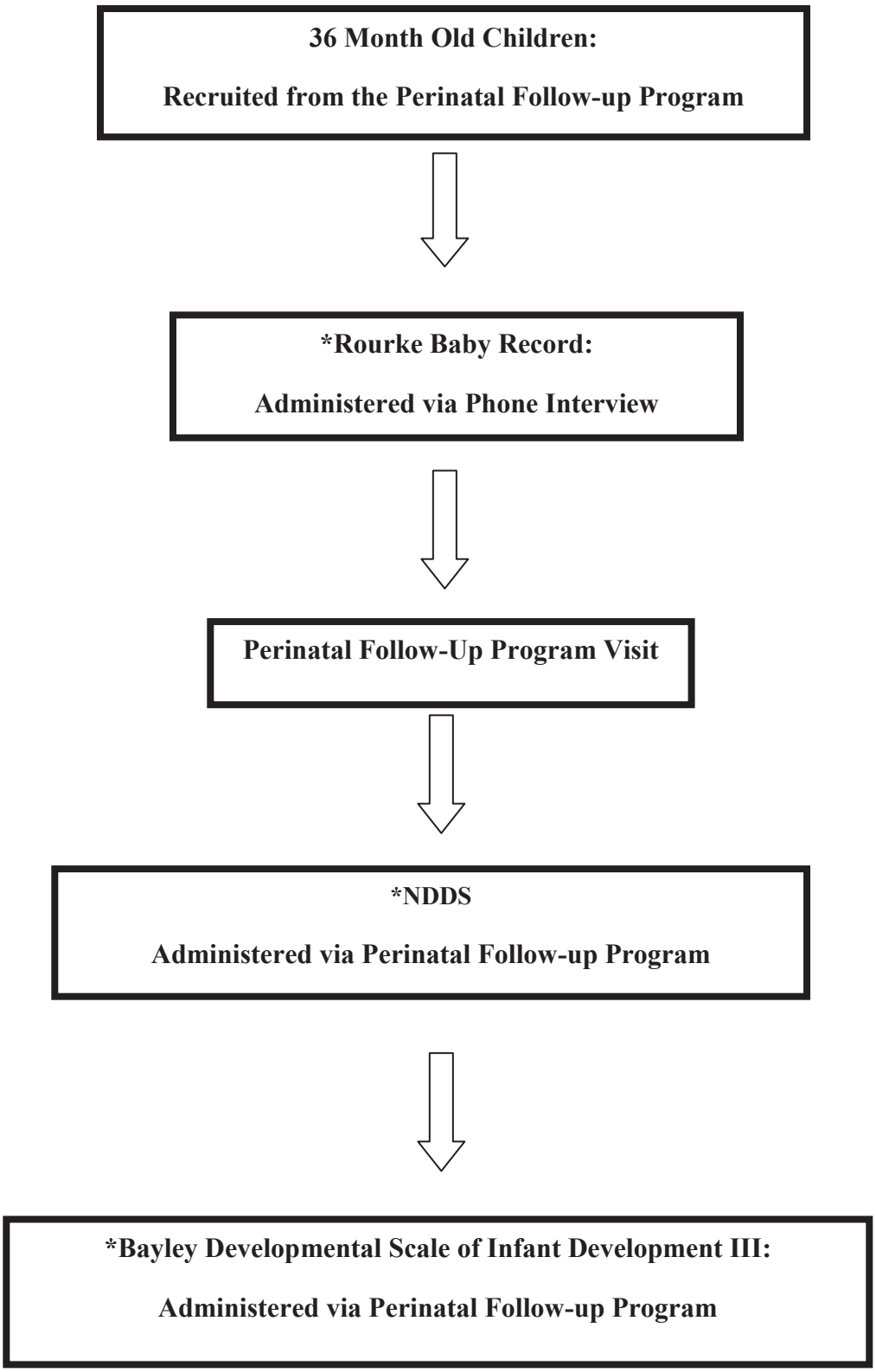
APPENDIX A: Standardized Assessments, Variables

BSITD-III	
1. Cognitive Scale	Continuous variable: questions vary dependant on child's ability
2. Motor Scale: a. Gross Motor b. Fine Motor	Continuous variable: questions vary dependant on child's ability
3. Language Scale: a. Receptive b. Expressive	Continuous variable: questions vary dependant on child's ability
4. Social Emotional Scale	Administered via parental questionnaire
5. Adaptive Behaviour Scale	Administered via parental questionnaire
Rourke: Thirty-six Month Developmental Assessment. ***"Does your child..."***	
1. Understands 2 and 3 step directions (e.g. "Pick up your hat and shoes and put them in the closet")	Yes/No
2. Uses sentences with 5 or more words	Yes/No
3. Walks up stairs using a handrail	Yes/No
4.Twists lids off jars and turns doorknobs	Yes/No
5. Shares some of the time	Yes/No
6. Plays make-believe games with actions and words (e.g. pretending to cool a meal, fix a car)	Yes/No
7. Turn pages one at a time	Yes/No
8. Listens to music or stories for 5-10 minutes	Yes/No
9. No parent/caregiver concerns	Yes/No

NDDS: Thirty-six Month Developmental Assessment: ***"By thirty-six months of age, does your child...."***	
1. Speak clearly enough to be understood most of the time by family?	Yes/No
2. Understand two-step directions (e.g. "Pick up your shoes and put them in the closet")?	Yes/No
3. Speak in two to five word sentences (e.g. "I go home now")?	Yes/No
4. Correctly say the words: my, home, pie, hop, bee, bib, no, man, one?	Yes/No
5. Understand and use some describing words like big, dirty, wet or hot?	Yes/No
6. Walk up the stairs using the handrail?	Yes/No
7. Stand on one foot briefly?	Yes/No
8. Throw a ball forward at least one metre (three feet)?	Yes/No
9. Twist lids off jars or turn knobs?	Yes/No
10. Turn the pages of a book one at a time?	Yes/No
11. Play make-believe games with actions or words?	Yes/No
12. Dress or undress with help?	Yes/No
13. Share some of the time (e.g. toys, books)?	Yes/No
14. Show affection with words and actions?	Yes/No
15. Play alongside others comfortably?	Yes/No
16. Cooperate with parent's request half of the time?	Yes/No
17. Listen to music or stories for 5 to 10 minutes with you?	Yes/No
18. Greet friends and familiar adults when reminded?	Yes/No

Bayley (2006); Rourke (2009); NDDS (2007)

APPENDIX B: Test Administration Procedures



APPENDIX C: Formulas Used for Assessment of Test Characteristics

Sensitivity: True Positive/ (True Positive + False Negative)

Specificity: True Negative/ (False Positive + True Negative)

Positive Predictive Value: True Positive/ (True Positive + False Positive)

Negative Predictive Value: True Negative/ (False Negative + True Negative)

False Positive Rate: False Positives / (False Positives + True Negatives)

False Negative Rate: False Negative / (False Negative + True Positive)

Likelihood Ratio: Sensitivity / False Positive Rate

Accuracy: (True Positive + True Negative) /Total Observations

Table 1.1: ***Descriptive Statistics of Participants***

	Mean	**Median**	**SD**	**Range**
Chronological Age at Assessment (months)	37.6	38.1	1.8	33.2-41.0
Gestational Age (weeks)	31.6	30.6	4.6	26.2-41.6
Birth Weight (grams)	1637.19	1350.0	878.21	690-4481
	Proportion			
Twin Gestation	41.94%			
Male Sex	54.84%			
Prematurity (birth prior to 31 weeks gestation)	54.84%			
Prematurity (birth prior to 37 weeks gestation)	80.65%			
Medical Complications at Birth				
Neurological Complications at Birth	41.94%			
Cardiac Complications	23.33%			
Respiratory Complications	61.29%			
Fetal Malnutrition	25.8%			

Table 1.2: ***Participant Response Rate***

	Raw Value	Proportion
Did not attend appointment	1	1.56%
Opt-out (verbal)	5	7.81%
Phone Disconnected	7	10.94%
Opt-out (mail response)	9	14.06%
No response by telephone	11	17.19%
Participated	31	48.44%
Total Participants Contacted	64	

Table 2.1: *Response Rate for each Question on the Rourke Baby Record: Thirty-Six Months, Developmental Section*

Does your Child...	Raw Value	Proportion of participants with a 'no' response (e.g., flagged positive) (%)
Understands 2 and 3 step directions (e.g. "Pick up your hat and shoes and put them in the closet")	1	3.2
Uses sentences with 5 or more words	6	19.4
Walks up stairs using a handrail	1	3.2
Twists lids off jars and turns doorknobs	1	3.2
Shares some of the time	1	3.2
Plays make-believe games with actions and words (e.g. pretending to cool a meal, fix a car)	4	12.9
Turn pages one at a time	1	3.2
Listens to music or stories for 5-10 minutes	3	9.7
No parent/caregiver concerns	5	16.1

Table 2.2: *Response Rate for each Question on the NDDS-Thirty-Six Months*

Does your Child...	Raw Value	Proportion of participants with a 'no' response (e.g., flagged positive) (%)
Speak clearly enough to be understood most of the time by family?	4	12.9
Understand two-step directions (e.g. "Pick up your shoes and put them in the closet")?	0	0
Speak in two to five word sentences (e.g. "I go home now")?	3	9.7
Correctly say the words: my, home, pie, hop, bee, bib, no, man, one?	3	9.7
Understand and use some describing words like big, dirty, wet or hot?	0	0
Walk up the stairs using the handrail?	1	3.2
Stand on one foot briefly?	3	9.7
Throw a ball forward at least one metre (three feet)?	3	9.7
Twist lids off jars or turn knobs?	2	6.5
Turn the pages of a book one at a time?	1	3.2
Play make-believe games with actions or words?	3	9.7
Dress or undress with help?	4	12.9
Share some of the time (e.g. toys, books)?	0	0
Show affection with words and actions?	0	0
Play alongside others comfortably?	0	0
Cooperate with parent's request half of the time?	0	0
Listen to music or stories for 5 to 10 minutes with you?	0	0
Greet friends and familiar adults when reminded?	1	3.2

Table 3.0: ***Descriptive Analysis of BSITD-III Scores, Stratified by Scale***

Composite Scale	N	Mean	SD	Median	Range
Motor	28	96.50	13.35	98.50	46-121
Cognitive	31	95.32	1.04	100	55-110
Language	30	99.90	15.96	103	56-114

Table 4.0: ***Descriptive Analysis of BSITD-III Scores, Stratified by Subscale***

	Fine Motor	Gross Motor	Cognitive	Expressive Communication	Receptive Communication
Scale Score					
N	30	29	31	31	30
Mean	9.73	9.00	9.10	10.0	9.87
SD	2.41	2.58	2.10	3.17	2.53
Median	10	9	10	10	10
Range	1-14	1-15	1-12	2-15	3-13
Developmental Age					
N	31	28	31	31	29
Mean	36.61	34.04	33.23	36.74	35.76
SD	6.87	9.59	6.28	9.05	8.55
Median	38	35	35	41	38
Range	11-52	10-66	12-46	12-46	14-43

Table 5.1 i-vi ***Score Allocation of BSITD-III (≤85) to the Screening Tools at Various Flag Cut Points***

Table 5.1.i: Test Characteristics of the 1 Flag Cut Point Relative to the BSITD-III

Rourke:1 Flag Cut Point	BSITD-III +	BSITD-III -	Total
Rourke +	3	7	10
Rourke –	1	20	21
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	26%
Specificity:	74%	False Negative Rate:	25%
PPV:	30%	Likelihood Ratio:	2.90
NPV:	95%	Accuracy:	74%

Table 5.1.ii: Test Characteristics of the Rourke with the Omission of Question 9 at 1 Flag Cut Point Relative to the BSITD-III

Rourke: 1 Flag Cut Point, Omission of Question 9	BSITD-III +	BSITD-III -	Total
Rourke +	3	6	9
Rourke –	1	21	22
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	22%
Specificity:	78%	False Negative Rate:	25%
PPV:	33%	Likelihood Ratio:	3.41
NPV:	95%	Accuracy:	77%

Table 5.1.iii: Test Characteristics of the Rourke at 2 Flag Cut Points Relative to the BSITD-III

Rourke: 2 Flags Cut Point	BSITD-III +	BSITD-III -	Total
Rourke +	3	2	5
Rourke –	1	25	26
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	7%
Specificity:	93%	False Negative Rate:	25%
PPV:	60%	Likelihood Ratio:	10.71
NPV:	96%	Accuracy:	90%

Table 5.1.iv: Test Characteristics of the Rourke with the Omission of Question 9 at 2 Flag Cut Points Relative to the BSITD-III

Rourke: 2 Flags Cut Point, Omission of Question 9	BSITD-III +	BSITD-III -	Total
Rourke +	3	2	5
Rourke –	1	25	26
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	7%
Specificity:	93%	False Negative Rate:	25%
PPV:	60%	Likelihood Ratio:	10.71
NPV:	96%	Accuracy:	90%

Table 5.1v: Test Characteristics of the NDDS at 1 Flag Cut Point Relative to the BSITD-III

NDDS: 1 Flag Cut Point	BSITD-III +	BSITD-III -	Total
NDDS +	3	6	9
NDDS –	1	21	22
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	22%
Specificity:	78%	False Negative Rate:	25%
PPV:	33%	Likelihood Ratio:	3.41
NPV:	95%	Accuracy:	77%

Table 5.1.vi: Test Characteristics of the NDDS at 2 Flag Cut Points Relative to the BSITD-III

NDDS: 2 Flags Cut Point	BSITD-III +	BSITD-III -	Total
NDDS +	3	1	4
NDDS –	1	26	27
Total	4	27	31
Sensitivity:	75%	False Positive Rate:	4%
Specificity:	96%	False Negative Rate:	25%
PPV:	75%	Likelihood Ratio:	20.27
NPV:	96%	Accuracy:	94%

Table 5.2 i-vi: ***Score Allocation of BSITD-III (≤70) to the Screening Tools at Various Flag Cut Points***

Table 5.2.i: Test Characteristics of the Rourke at 1 Flag Cut Point Relative to the BSITD-III

Rourke: 1 Flag Cut Point	BSITD-III +	BSITD-III -	Total
Rourke +	2	8	10
Rourke –	0	21	21
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	28%
Specificity:	72%	False Negative Rate:	0%
PPV:	20%	Likelihood Ratio:	3.62
NPV:	100%	Accuracy:	74%

Table 5.2.ii: Test Characteristics of the Rourke with Question 9 Omitted, at 1 Flag Cut Point Relative to the BSITD-III

Rourke: 1 Flag Cut Point, Omission of Question 9	BSITD-III +	BSITD-III -	Total
Rourke +	2	7	9
Rourke –	0	22	22
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	24%
Specificity:	76%	False Negative Rate:	0%
PPV:	22%	Likelihood Ratio:	4.15
NPV:	100%	Accuracy:	77%

Table 5.2.iii: Test Characteristics of the Rourke at 2 Flag Cut Points Relative to the BSITD-III

Rourke: 2 Flags Cut Point	BSITD-III +	BSITD-III -	Total
Rourke +	2	3	5
Rourke –	0	26	26
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	10%
Specificity:	90%	False Negative Rate:	0%
PPV:	40%	Likelihood Ratio:	10.34
NPV:	100%	Accuracy:	90%

Table 5.2.iv: Test Characteristics of the Rourke with Question 9 Omitted, at 2 Flag Cut Points Relative to the BSITD-III

Rourke: 2 Flags Cut Point, Omission of Question 9	BSITD-III +	BSITD-III -	Total
Rourke +	2	3	5
Rourke –	0	26	26
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	10%
Specificity:	90%	False Negative Rate:	0%
PPV:	40%	Likelihood Ratio:	10.34
NPV:	100%	Accuracy:	90%

Table 5.2.v: Test Characteristics of the NDDS at 1 Flag Cut Point Relative to the BSITD-III

NDDS: 1 Flag Cut Point	BSITD-III +	BSITD-III -	Total
NDDS +	2	7	9
NDDS –	0	22	22
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	24%
Specificity:	76%	False Negative Rate:	0%
PPV:	22%	Likelihood Ratio:	4.15
NPV:	100%	Accuracy:	77%

Table 5.2.vi: Test Characteristics of the NDDS at 2 Flag Cut Points Relative to the BSITD-III

NDDS: 2 Flag Cut Points	BSITD-III +	BSITD-III -	Total
NDDS +	2	2	4
NDDS –	0	27	27
Total	2	29	31
Sensitivity:	100%	False Positive Rate:	7%
Specificity:	93%	False Negative Rate:	0%
PPV:	50%	Likelihood Ratio:	14.50
NPV:	100%	Accuracy:	94%

Table 6.0: ***Crude and Adjusted Logistic Regression for Rourke and NDDS at one standard deviation below the mean on the BSITD-III***

	N	Crude Analysis			Adjusted Analysis**		
		OR	**95% CI**	**P**	**OR**	**95% CI**	**P**
Rourke 1 Flag	10	8.57	0.76-96.52	0.08	10.61	0.74-152.57	0.08
Rourke 2 Flags	5	37.5	2.56-548.36	0.008	*	*	*
NDDS 1 Flag	9	10.5	0.92-120.26	0.06	23.84	0.73-774.95	0.07
NDDS 2 Flags	4	78.0	3.81->999.99	0.005	*	*	*

Referent Category = No flags of concern

*Not possible to complete adjusted analysis of this models due to low cell size

**Adjusted for twin gestation and gestational age

Table 7.0: ***Pearson Correlation Coefficients for Rourke and NDDS at each flag cut point and total score, relative to subscales of the BSITD-III.***

Value	BSITD-III: 1 SD Below the Mean	BSITD-III: 2 SD Below the Mean
Rourke: one flag	0.35 (p=0.05)	0.38 (p=0.03)
Rourke: two flags	0.62 (p=0.0002)	0.60 (p=0.0004)
Rourke: > two flags	0.71 (p<0.0001)	0.68 (p<0.0001)
NDDS: one flag	0.39 (p=0.03)	0.41 (p=0.02)
NDDS: two flags	0.71 (p<0.0001)	0.68 (p<.0001)
NDDS: > two flags	0.85 (p<0.0001)	0.81 (p<0.0001)