USE OF THE REVISED BAYLEY SCALES OF INFANT DEVELOPMENT WITH HIGH-RISK INFANTS: EXPLORATION OF CHANGES IN SCORES AND RELATIONSHIPS BETWEEN RISK VARIABLES AND PERFORMANCE

By

KAREN KOSECK

B. Sc. (OT), The University of Western Ontario, 1991

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

MASTER OF SCIENCE

In

THE FACULTY OF GRADUATE STUDIES

(School of Rehabilitation Sciences)

We accept this thesis as conforming
To the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2000

Department of Rehabilitation Sciences

The University of British Columbia
Vancouver, Canada

Date Aug. 20, 2000

Abstract

Cognitive and mental performance of 53 high-risk infants (34 prenatally exposed to drugs, and 19 born prematurely and/or with significant perinatal medical concerns) was measured using the Revised Bayley Scales of Infant Development (BSID-II) at less than 12 months, and at around 18 months of age. A retrospective chart review was used to collect BSID-II data and information on biological and environmental risk variables. The high-risk infants scored lower on the Mental and Motor Scales at both testing times compared to the norms ($p<.01$). Cognitive performance decreased over time if extrapolated scores were used in the analysis ($p<.01$). Infants born prematurely and/or with significant perinatal medical concerns had better overall motor performance compared to infants prenatally exposed to drugs ($p<.01$). Their motor scores increased, while the scores obtained by infants prenatally exposed to drugs decreased, regardless of whether or not extrapolated scores were used ($p<.01$). There were a few, isolated instances of linear relationships between risk variables, and cognitive and motor scores. The variance explained by regression equations tended to be low. Use of performance classifications versus extrapolated scores was the preferred method to include very low scores in analyses. Cumulative measures of environmental and biological risks could be better predictors of developmental outcomes than use of single variables.

TABLE OF CONTENTS

LIST OF TABLES

Acknowledgements

I would like to express tremendous gratitude towards my thesis committee: Dr. Susan R. Harris, Dr. Robert W. Schutz, and Dr. Anne Carswell. Your unfaltering mentorship, patience, encouragement and support were essential for the completion of this thesis.

A special thanks to Dr. Susan Harris for sharing the original data from her study, and providing the idea for this study. Her wisdom, warmth, and sense of humour were appreciated all along. The positive feedback she provided at all stages of the thesis helped keep me motivated. Her strength and determination through difficult times helped me grow professionally, and personally. I feel blessed to have had the opportunity to work with her.

Dr. Robert Schutz's expertise in statistics made it possible for me to analyze and interpret the data. His ability to explain complicated analyses in such a way that I could understand made him a gifted teacher. He could always strike the perfect balance to challenge me without frustration. I am proud of what I have learnt with his help.

Dr. Anne Carswell's solid advice helped shape the direction of this thesis. I appreciated her extra assistance during the times that she was the co-advisor. Her advice and support regarding my parental leave was appreciated by my family and myself.

I would also like to acknowledge Dr. Jon Berkowitz for helping me to sort out the data from the original study. His time and effort will not be forgotten.

I would like to express my gratitude for the efficient and friendly service provided by the Medical Records staff at British Columbia's Children's Hospital and Sunny Hill Health Centre for Children.

My employers and colleagues from the Children's Services Resource Team were supportive of my four-year academic adventure. Thank-you for encouraging me, and giving me with leave to complete my schooling.

I am extremely grateful for the scholarship provided by the Canadian Occupational Therapy Foundation. Your financial and professional support were vital.

A special thanks goes out to the children, families, and therapists who participated in the original study. Without their participation this study would not have been possible.

I would especially like to thank Jay, a devoted husband and father. His extra time spent in running the household and looking after our daughter gave me the time to do this. His encouragement and pep talks gave me the will.

Hugs and kisses to Megan who was born in the middle of this. Thanks for reminding me to enjoy life and spend some time playing Pooh Bear on the "pewter".

CHAPTER 1

Introduction

Statement of the Research Problem

Accurate measurement and prediction of infant development is required for early identification, and subsequent timely intervention, of infants at high risk for developmental delays. Occupational therapists and physical therapists, as members of a multidisciplinary team, frequently use standardized norm-referenced tests of infant development as part of a diagnostic battery for early identification of infants with a developmental delay.

The Bayley Scales of Infant Development, or BSID, (Bayley, 1969) have been described as the most widely used measure of cognitive and motor development in infancy (Aylward, Pfeiffer, Wright, & Verhulst, 1989; Cherny et al., 1994; Gross, Slagle, D'Eugenio, & Mettelman, 1992). However, it has been suggested that the BSID may have several weaknesses. Although useful in the identification of infants with gross developmental delays, it may not be useful for identifying the specific cognitive and motor deficits associated with some high risk populations, such as infants born prematurely (Aylward et al., 1989) and infants prenatally exposed to alcohol and/or other drugs (Chasnoff, Griffith, Freier, & Murray, 1992). Others have suggested that the BSID is unable to accurately measure mental ability because the Mental Scale contains many items with a motoric or sensory component (Bornstein & Sigman, 1986; Brooks-Gunn, Klebanov, Liaw, & Spiker, 1993). Lastly, it has been recommended that the BSID norms need to be revised (Campbell, Siegel, Parr, & Ramey,1986; Gross et al., 1992;

Richardson, Day, & Goldschmidt, 1995). Recent use of the test appears to result in inflated test scores, which are thought to be due to a true change in performance of young children since the test's development in 1969.

The Bayley Scales of Infant Development-Second Edition, or BSID-II (Bayley, 1993), represent a long-awaited revision and restandardization of the original Bayley Scales. The scales were revised in order to update normative data, expand the age range, improve content coverage, update materials, conduct reliability and validity studies, and improve clinical utility (Bayley, 1993). Although some of the concerns with the original Bayley Scales were addressed in the revised test, more evidence is needed in order to demonstrate that the BSID-II is able to identify special populations of high-risk infants for whom the test is commonly used.

Since its publication in 1993, there has been very little published on the BSID-II. Eight papers were descriptive reviews or commentaries (Flanagan & Alfonso, 1995; Gauthier, Bauer, Messinger, & Closius, 1999; Koseck, 1999; Matula & Aylward, 1997; Mayes, 1997; Nellis and Gridley, 1994; Ross & Lawson, 1997; Washington, Scott, Johnson, Wendel & Hay, 1998), two were reports with either extrapolated scores (Robinson & Mervis, 1996) or intrapolated scores (Lindsey & Brouwers, 1999), and one was a factor analysis (Thompson, Wasserman, & Matula, 1996). Three studies administered the BSID-II to typically-developing children (Levy-Shiff, Dimitrovsky, Shulman & Har-Even, 1998; Saudino et al., 1998; Tasbihsazan, Nettlebeck, & Kirby, 1997). Only three studies have been published which used the BSID-II with infants prenatally exposed to alcohol and/or other drugs (Alessandri, Bendersky, & Lewis, 1998; Cosden, Peerson, & Elliot, 1997; Heffelfinger, Craft, & Shyken, 1997). Five studies

(Case-Smith, Butcher, & Reed, 1998; Costarides & Shulman, 1998; Doig, Macias, Saylor, Craver, & Ingram, 1999; Goldstein, Fogle, Wieber, & O'Shea, 1995; Macias et al., 1998) and one abstract (Mattia & deRegnier, 1998) used the BSID-II with infants born prematurely and/or with significant perinatal medical concerns.

Because the rate of development in infancy is considered to be unstable, variable and changing (Aylward, Gustafson, Verhulst, & Colliver, 1987; Cole & Harris, 1992; Coryell, Provost, Wilhelm, & Campbell, 1989; Dunst & Rheingrover, 1981), sequential or serial assessments of infant development have been recommended (Campbell et al., 1986; Darrah, Redfern, Maguire, Beaulne, & Watt, 1998; Piper, Darrah, Pinnell, Watt, & Byrne, 1991). Only one study (Alessandri et al., 1998) and one abstract (Mattia & deRegnier, 1998) have collected BSID-II test scores on more than one occasion.

Alessandri and colleagues (1998) and Mattia and deRegnier (1998) explored potential relationships between risk variables and performance on the Bayley-II. Understanding of these relationships (i.e. factors which influence development) could contribute to a theoretical framework of development, which in turn could assist clinicians in planning interventions aimed at preventing or reducing developmental delays in infants at risk. For example, are there environmental risk variables that could be reduced through therapeutic interventions in the home? Are there biologic risk variables that could be reduced through public health education?

Overall, further research is needed to assess the clinical validity of the BSID-II scores, as well as the need for sequential assessments, before clinicians can adopt it as the new "gold standard" test. Exploration of biologic and environmental variables that place

infants at risk for developmental delays could be beneficial in formulating and refining developmental models and health care prevention initiatives.

Purposes of the Study

This study aimed to provide new knowledge, which will help enable clinicians to make informed decisions regarding effective and appropriate use of the BSID-II with populations of infants at high-risk for developmental delays. Specifically, this study has analyzed data which were collected on 53 high-risk infants (Harris, 1995). The sample included 34 infants who were prenatally exposed to alcohol and/or other drugs, and 19 non-exposed infants with significant perinatal medical factors that put them at risk for developmental delays. Thirteen of these nineteen infants were born prematurely (i.e. gestational age of 37 weeks or less). BSID-II Mental and Motor Scale test scores were collected on two occasions; once during the first year of life, and once at approximately 18 months of age.

The purposes of the study were (1) to compare performance on the BSID-II Mental and Motor Scales between the sample of high-risk infants and test norms, (2) to describe changes in test scores over time for the entire sample and for each high-risk subgroup (i.e. prenatally exposed to alcohol and/or other drugs, and infants born prematurely and/or with significant perinatal medical concerns), and (3) to explore relationships between biologic and environmental variables on early (i.e. before 12 months of age) and later (i.e. at around 18 months of age) BSID-II outcomes from high-risk infants.

Definitions

The following section defines terms, as they will be used throughout the thesis:

Biologic risk variables: perinatal medical complications thought to have a biologic or genetic basis, which are believed to place an infant at risk for a developmental delay (e.g. low birthweight, intraventricular hemorrhage, seizures, prenatal alcohol/drug exposure).

BSID-II raw score: non-standardized score obtained by adding the number of passed or credited items between basal and ceiling levels of the Bayley Scales of Infant Development-II.

BSID-II item set: specifies the group of items to administer to a child based on the child's age.

Corrected age: age used for assessing developmental status of infants born prematurely, calculated by subtracting the amount of time the infant was premature from the infant's chronological age.

Environmental risk variables: factors in the infant's environment which are thought to place an infant at risk for a developmental delay (e.g. low socioeconomic status, high stress in the household, maternal depression).

High-risk infant: infants considered to be at risk for having a developmental delay due to pre- or perinatal risk factors (e.g. prenatal exposure to alcohol and/or other drugs, prematurity, or other medical concerns).

Internal consistency: refers to the consistency of measurement, itself, or how well the test items measure the same variable.

Inter-rater reliability: refers to the consistency of test scores when they are determined by different examiners.

Mental Developmental Index (MDI): Standardized score for the BSID-II Mental Scale (1993) with mean of 100 and standard deviation of 15. Also, the standardized score for the BSID Mental Scale (1969) with mean of 100 and standard deviation of 16.

Performance classifications: categories of performance based on the distance in standard deviations of the test score from the normative mean.

Psychomotor Developmental Index Score (PDI): Standardized score for the BSID-II Motor Scale (1993) with mean of 100 and standard deviation of 15. Also, the standardized score for the BSID Motor Scale (1969) with mean of 100 and standard deviation of 16.

Test-retest reliability: measurement of the stability of test scores over time by comparing examinees' performance on two or more separate administrations of the test.

CHAPTER 2

Literature Review

Measuring Development

Measurement of human development is extremely challenging because there is no universally accepted theory of development. There are many different theories of development, each defining the construct of development somewhat differently (Favell, Miller & Miller,1993; Thelen, 1995). Because different developmental assessments could be measuring different constructs, making comparisons between test scores from different instruments may not be valid. Therefore, it is in the best interest of clinicians to use the same measurement instrument when comparing performance among different infants, and when evaluating progress within the same infant.

Uses of the BSID-II include the identification of children who are developmentally delayed and the evaluation of progress or developmental change (Bayley, 1993). Evaluation of developmental status and developmental change both rely on the reliability of the test scores obtained.

Reliability of BSID-II Test Scores

Reliability refers to the stability and consistency of test scores. Sources in measurement error can arise from test characteristics, examiners' testing ability, and examinees' performance variability. Bayley (1993) claimed that the BSID-II is a very reliable instrument. However, in a critique of preschool intelligence tests by Flanagan and Alfonso (1995), the BSID-II Mental Scale reliability was evaluated for the 30 to 42-month age range, and rated as inadequate. Unfortunately, their review did not include a

critique of the Motor Scale, nor a critique of reliability of test scores from children of younger ages (i.e. 1 to 29 months of age).

In an evaluation of the psychometric properties of the BSID-II, different types of reliability coefficients presented in the manual were critiqued (Koseck, 1999). Koseck reported that most of the BSID-II reliability coefficients met the set criterion of $r \geq 0.80$. A few internal consistency coefficients for MDI and PDI scores, as well as a few test-retest and interrater coefficients for PDI scores, did not meet this criterion. No intra-rater reliability studies were presented, or consistency of test scores obtained by the same rater for the same testing session. Cunningham-Amundson and Crowe (1993) suggested that examiners could increase the reliability of test scores through the use of the standard error of measurement (SEM). Confidence intervals based on SEM are easily derived from index scores using tables presented in the BSID-II manual.

Potential Scoring Problems

Lack of clear administration and scoring instructions can reduce the reliability of test scores. To date, there have been five commentaries on potential scoring problems associated with administering the BSID-II to clinical populations (Gauthier et al., 1999; Mayes, 1997; Nellis & Gridley, 1994; Ross & Lawson, 1997; Washington et al., 1998). Nellis and Gridley (1994) commented on a possible discrepancy in test scores for children with developmental delays, depending on the item set used to begin testing. It was cautioned that if a child was started on an item set below his/her chronological age because of a suspected developmental delay, the score could be significantly lower than if the child was started on his/her chronological item set, and then was presented items in descending order until a basal level was established. The basal level established through

working in descending order could be higher than the one obtained when working in ascending order. Due to the liberal BSID-II basal and ceiling rules, there is a possibility that the same child could meet basal and ceiling criteria using different item sets.

Mayes (1997) had similar concerns regarding the latitude in determining where to begin testing children with developmental delays. Mayes administered the entire BSID-II Mental Scale to a sample of 32 children typically referred to developmental testing (e.g. children with brain injury, autism, cerebral palsy, metabolic disorders etc.). Children were tested downwards in item sets until the child passed all the items, and upwards until the child failed all the items in an item set. Results indicated that the higher the age associated with the starting item set, the higher the obtained score. In other words, there is a potential for different examiners to obtain different scores for the same child depending on which item set was chosen to begin testing.

In a study by Gauthier and colleagues (1999), the BSID-II Mental Scale was administered to 78 twelve-month-old infants who were prenatally exposed to cocaine. One examiner began testing with the 12-month item set, and another examiner administered 10 additional items that completed the 11- and 13-month item sets. Ninety-four percent of the sample met basal and ceiling criteria for all three item sets. Scores obtained using the 11-month item set were significantly lower than those obtained using the 12-month item set, which were significantly lower than those obtained using the 13-month item set ($p<.0001$). Furthermore, twice as many infants were classified as at-risk/delayed (i.e. MDI<85, or greater then one standard deviation below the normative mean) when scored using the 11-month versus the 12-month item set (37% and 18% respectively). When the 13-month item set was used, infants classified as at-risk/delayed

were virtually eliminated (3%). To reduce inconsistencies in test administration, the authors suggested that all full-term infants be started on the item set corresponding to their chronological age and tested downwards, if needed, to establish a basal level. Whether to use corrected or chronological age as a starting point for infants born prematurely still needs to be decided and used consistently by examiners.

Washington and colleagues (1998) presented four case studies where problems arose with the BSID-II item-set format when children with atypical development were tested. In one case, a child with a substantial gross-motor delay received a misleadingly high PDI score because of the child's fine motor ability. In the other three cases, it was possible to obtain more than one score depending on the item set(s) used in test administration. Therefore, the authors suggested that caution be used in clinical practice when the BSID-II is administered to children who have substantial delays or to children who have uneven developmental profiles.

Ross and Lawson (1997) expressed concern regarding which item set should be used as a starting point when administering the BSID-II to children born prematurely. MDI scores were calculated using chronological and corrected ages for 100 very low birthweight premature infants. As expected, Ross and Lawson found that scores were significantly lower when testing started on the item set corresponding to the infant's corrected versus chronological age. Furthermore, they commented that the BSID-II manual did not provide a standard age of when to stop correcting for prematurity.

Matula and Aylward (1997) responded to some of Ross and Lawson's concerns. They recommended using the child's chronological-age item set to begin testing, because that should be appropriate for examinees performing within 1.5 standard deviations of

what is expected for their age. Secondly, they advised that correcting for prematurity should by decided by the examiner, and would depend upon the purpose of testing, and the age and length of the examinee's gestation. Whichever the examiner's decision, it was advised that the norm table used to convert raw to standard scores should always correspond to the starting item set.

Similarly, the BSID-II manual clearly directs examiners to administer the item set corresponding to the child's chronological age with a footnote suggesting that "when testing a premature child under the age of 2 years, you may want to begin testing with the item set appropriate for the child's corrected age" (Bayley, 1993, p. 41). It appears that the starting item set is left to the discretion of the examiner, but that the 2-year mark should be used as the age at which examiners stop correcting for prematurity.

In summary, the reliability of BSID-II test scores could be further reduced than that reported in the manual when testing clinical populations, due to flexibility in test starting point as determined by individual examiners. If the basal and ceiling criteria were defined in a way that the same child could meet these criteria for only one particular sequence of item sets, this problem in scoring could be alleviated.

Early Identification of Infants With Developmental Delays

Measurement of developmental status is extremely challenging and so, also, is the accurate identification of children who are developmentally delayed. There is no universally accepted and standardized definition of developmental delay. Instead, standardized test scores are usually interpreted in relation to the normative mean. The difference in test score from the mean that is considered clinically significant has not been firmly established. However, the BSID-II manual contains guidelines for converting

test scores into performance classifications. These BSID-II performance classifications are based on the distance in standard deviations of MDI or PDI scores from the normative mean. Classifications are as follows: 115 and above is classified as accelerated performance, 85-114 as within normal limits, 70-84 as mildly delayed performance, and 69 and below as significantly delayed performance.

Performance classifications could be useful to clinicians in the identification of infants with developmental delays (i.e. classified as mildly or significantly delayed performance). However, as the BSID-II is a newer testing instrument, and as validity evidence is cumulative in nature (Dunn, 1989), caution should be used when using BSID-II performance classifications in interpretation of test scores.

Flanagan and Alfonso (1995) suggested that much of the validity research on the BSID is applicable to the BSID-II. Nellis and Gridley (1994) argued that since the correlations between the BSID and the BSID-II MDI and PDI scores reported in the BSID-II manual for 200 typically developing children were only moderate ($r = 0.62$ and $r = 0.63$ respectively), extreme caution should be used not to overgeneralize results from BSID studies. D.J. Goldstein and colleagues (1995) obtained much higher correlations between the BSID and BSID-II test scores for 49 preterm infants (i.e. $r = 0.95$ for both MDI and PDI scores). However, because their study used a smaller sample whose performance was more variable, and because the time interval between administration of the BSID and BSID-II was shorter, larger correlations could be expected.

Although it is questionable whether studies using the BSID can help to validate the BSID-II, a brief summary of BSID study results will be presented followed by a more in-depth review of studies using the BSID-II. Because the original Bayley Scales appear

to be based on outdated norms, BSID scores are expected to be higher than BSID-II scores.

The sample of high-risk infants in the present study consisted mainly of infants prenatally exposed to alcohol and/or other drugs, and infants with significant perinatal medical complications, most of whom were born prematurely. The studies reviewed in the following sections will be divided into two high-risk subgroups that most accurately reflect the participants in the present study: infants exposed to drugs and infants born prematurely.

Infants Prenatally Exposed to Drugs

Related Studies Using the Original Bayley Scales

Nineteen studies were reviewed which used the BSID to measure developmental outcomes of infants prenatally exposed to alcohol and/or other drugs (refer to Appendix A). Of the studies which did not use control groups to compare means, most obtained MDI and PDI group means that were average or better, that is, means of 95 or higher (Howard, Beckwith, Espinosa, & Tyler, 1995; Jacobson et al., 1993; O'Connor, Brill, & Sigman, 1986; O'Connor, Sigman, & Kasari, 1993; Richardson et al., 1995; Seagull et al., 1996; Streissguth, Barr, Martin, & Herman, 1980).

In a study by Fried and Watkinson (1988), group means from a sample of alcohol- and marijuana-exposed infants were compared to the normative mean. The sample obtained significantly higher MDI means at 12 and 24 months of age. The PDI means did not differ significantly from the normative mean. In a study by Mellins, Levenson, Zawardzki, Kairam, and Weston (1994) the prenatal drug-exposed group obtained a mean MDI and PDI over one-half of a SD below the normative mean.

Due to the outdated BSID norms, it was strongly recommended by Gross and colleagues (1992) that a matched control group be used when interpreting developmental performance using the test. In all studies where performance of infants exposed to alcohol was compared to control group performance, the alcohol-exposed groups obtained significantly lower MDI (O'Connor et al., 1986, 1993) and PDI scores (Golden, Sokol, Kuhnert, & Bottoms, 1982). In a longitudinal study by Chasnoff and colleagues (1992), the group exposed to alcohol and/or marijuana obtained significantly lower MDI and PDI scores at 6 months of age, and lower PDI scores at 12 and 18 months of age compared to the control group.

Results of studies comparing opiate-exposed (i.e. heroin or methadone) children to non-exposed control children were inconsistent. In a study by Rosen and Johnson (1982), the children in the methadone-exposed group obtained significantly lower MDI and PDI scores compared to the control group at 12 and 18 months, but not at 6 months of age. In two other studies, there were instances where MDI or PDI scores were significantly lower than those for the control group (Chasnoff, Burns, Burns, & Schnoll, 1986; Wilson, 1989). Because these instances were few and isolated, the differences in scores were thought to be clinically insignificant. Lastly, in a study by Hans (1989), no significant differences between methadone-exposed and control group MDI scores were found, until the groups were further divided into high and low socioeconomic status (SES). Then, low-SES methadone-exposed children performed more poorly than low-SES control children did.

Mixed results were also obtained when comparing scores of cocaine/polydrug-exposed children to control groups. Studies by Chasnoff and colleagues (1992), and Hurt

and colleagues (1995) found only isolated and inconsistent differences in BSID test performance between cocaine-exposed and control groups. In a study by Johnson, Seikel, Madison, Foose, and Rinard (1997), the control group outperformed the cocaine-exposed group on the Mental Scale. In a study by Billman, Nemeth, Hiemler, and Sasidharan (1996) and where groups were stratified by race (i.e. Black and White), the Black cocaine-exposed group obtained higher PDI scores than the Black control group. The authors suggested that black infants could have genetically increased susceptibility to certain CNS-maturing influences, with cocaine being one of them.

Two studies (Chasnoff et al, 1992; Golden et al., 1982) further compared results between exposed and control groups by the number of individuals whose scores fell below one SD from the normative mean (i.e. the number of subjects whose performance was classified as either mildly or significantly delayed). In both of these studies, more children exposed to alcohol and/or other drugs scored below one SD compared to non-exposed children. Due to outdated means, it is possible that the number of individuals identified in these studies as having delays had significant delays, and that children with mild delays were not identified.

In the studies reviewed using the BSID to assess developmental levels of children exposed to alcohol and/or other drugs, use of control groups for comparison of means appeared to be essential for identification purposes as BSID scores are likely to be inflated. Although, results using control groups were inconsistent, when differences in scores did exist, control groups tended to outperform exposed groups. Performance classifications appeared to be useful for the identification of individuals with

developmental delays; however, it is possible only children with significant delays were identified.

Related Studies Using the Revised Bayley Scales

A clinical validity study investigating the performance of 137 children prenatally exposed to drugs is presented in the BSID-II manual. Most of the children had been exposed to multiple drugs including alcohol, marijuana, cocaine, heroin, amphetamines, and/or nicotine. These children obtained a mean MDI score of 90.8 with a SD of 16.2, and a mean PDI score of 96.3 with a SD of 20.0. It was concluded that children prenatally exposed to drugs (1) displayed slightly greater variability in performance on the Motor Scale, (2) performed relatively more poorly on the Mental Scale compared to the Motor Scale, and (3) performed in the average to low average range in relation to the normative mean (Bayley, 1993). Although the mean MDI score was more than one-half a SD below the normative mean, it was not reported if the differences in scores were statistically and/or clinically significant.

Alessandri and colleagues (1998) used the BSID-II Mental and Motor Scales to examine the cognitive and motor functioning in 112 infants, of whom 15 were exposed to high levels of cocaine prenatally, 19 were exposed to low levels of cocaine, and 78 were non-exposed. At eight months of age, once neonatal and environmental risks were accounted for, there were no significant differences in mean MDI scores (means ranged from 91.30 to 94.59) or mean PDI scores (means ranged from 87.33 to 95.20) among the three groups. At eighteen month of age, the mean MDI scores were 79.05 (SD=10.21) for the high cocaine-exposed group, 86.59 (SD=9.76) for the low cocaine-exposed group, and 83.09 (SD=12.82) for the non-exposed group. The non-exposed infants obtained

higher MDI scores relative to the high cocaine-exposed infants ($p<.05$). Eighteen-month PDI scores were not obtained. The authors concluded that at 8 months of age, cognitive and motor functioning in cocaine-exposed infants were comparable to infants of similar backgrounds and to the scores obtained in the BSID-II manual for "at risk" populations. Eighteen-month MDI means were at least one-half SD below the normative mean for exposed and non-exposed infants.

In a study by Heffelfinger and colleagues (1997), the BSID-II Mental Scale was administered to 31 children aged 8 to 40 months, 17 children with prenatal exposure to cocaine, and 14 children without drug exposure matched on age, prematurity, birthweight, maternal education, ethnicity, and gender. The mean MDI score of 86.64 (SD=12.0) for the exposed group was significantly lower than the mean MDI score of 99.47 (SD=12.18) for the control group (p-value not reported). It was concluded that cocaine-exposed children were delayed in cognitive development.

As part of a study by Gauthier and colleagues (1999), the BSID-II Mental Scale was administered to 78 infants exposed to cocaine at 12 months of age. The mean MDI obtained was 95.03 (SD= 9.71). Although not tested for significance, the mean score appears to be within normal limits.

Cosden and colleagues (1997) assessed 80 infants of mothers who were enrolled in a treatment facility for drug addiction. Fifty infants were administered the original Bayley Mental and Motor Scales and thirty were administered the revised edition. Test scores from both versions of the Bayley were combined in the analyses. However, the BSID-II manual presents only a moderate correlation between the BSID and BSID-II (MDI r=0.62 and PDI r= 0.63) for a sample of typically-developing children aged 1 to 42

months. A comparison of mean MDI and PDI scores showed approximately a 12-point drop in MDI and a 7-point drop in PDI scores from the BSID to the BSID-II (Bayley, 1993).

A study by Tasbihsazan and colleagues (1997) comparing the two versions of the Bayley Mental Scale for a sample of 97 typically-developing children obtained results similar to those presented in the BSID-II manual. Although the correlations between the BSID and BSID-II test scores were higher (ranging from $r = 0.84$ to 0.93), the authors felt that this was due to the narrower age range (18 to 27 months) that was used. Drops in MDI scores on the BSID-II ranged from 4 to 35 points.

In a study by D.J. Goldstein and colleagues (1995) of 49 high-risk preterm infants aged 12 to 22 months, a high correlation was found between the original and revised Bayley Scales ($r = 0.95$ for MDI and PDI) with an average drop of 7.3 index points on the BSID-II Mental Scale and 9.3 points on the Motor Scale. Mean scores were significantly lower on the revised Bayley Scales ($p < .001$ for both MDI and PDI scores).

Because correlations between the BSID and the BSID-II have been variable, it is not clear whether the two versions of the test are measuring identical constructs. Furthermore, with updated norms on the second version of the test, infants can be expected to obtain lower scores. Combining BSID and BSID-II scores in data analyses does not appear to be appropriate. Therefore, the results from the study by Cosden and colleagues (1998) can not be accurately interpreted.

With only a limited number of studies available using the BSID-II, extreme caution must be used in drawing any conclusions regarding the cognitive and motor development of infants prenatally exposed to drugs as measured by the BSID-II. Only

two studies collected BSID-II Motor Scale scores (Alessandri et al.,1998; Bayley, 1993). In one of these studies (Bayley, 1993) infants prenatally exposed to drugs obtained better motor compared to cognitive test scores. The literature review revealed the following trends in cognitive development of infants prenatally exposed to drugs: (1) development of cognitive skills may be delayed, but not identified by the BSID-II until later in life (at 18 months-of-age), and (2) cognitive scores may decrease over time, especially for infants with higher biologic and environmental risk.

Infants Born Prematurely

Related Studies Using the Original Bayley Scales

Twenty-four studies were reviewed where the BSID was used to measure developmental outcomes in preterm infants and children (refer to Appendix B). The majority of the studies reviewed, that did not use a control group, obtained means which were one-half a SD below the normative mean or better, that is, MDI and PDI means of 92.5 or higher (Bendersky & Lewis, 1994; Feingold, 1994; Gennaro & Stringer, 1991; Lipkin & Altshuler, 1994). Three studies obtained MDI means of 92.5 or higher but some PDI means below 92.5 (Brazy, Eckerman, Oehler, Goldstein, & O'Rand, 1991; Gross et al., 1992; Thompson, et al., 1994). A study by Brazy, Goldstein, Oehler, Gustafson, and Thompson, (1993) had some cases where both MDI and PDI means were more than one-half a SD below the norm.

Because the BSID is believed to have outdated norms, mean test scores are probably inflated; therefore, comparison of preterm infant performance to that of control groups could be more beneficial (Gross et al., 1992). Group assignment varied between the studies reviewed (refer to Appendix B). Some studies assigned infants to different

groups based on the degree of biologic or medical risk. Lewis and Bendersky (1989) found that preterm infants with severe intraventricular hemorrhage, or IVH, (i.e. Grades III and IV) obtained significantly lower MDI and PDI scores compared to preterm infants with mild IVH (i.e. Grades I and II). In a study by Ross, Tesman, Auld, and Nass (1992), preterm infants with subependymal and mild IVH (S/IVH) obtained significantly lower MDI, but not PDI scores, compared to preterm infants without S/IVH and to full-term infants. In a later study by Ross, Boatright, Auld, and Nass (1996), no significant differences in MDI or PDI scores were found between the groups.

Anderson and colleagues (1996) assigned infants to high-risk preterm, low-risk preterm, and full-term groups. Level of biologic risk of preterm infants was determined by the number and severity of perinatal medical complications. At six months of age, full-term infants obtained significantly higher MDI and PDI scores compared to low-risk preterm infants, who obtained higher scores compared to high-risk preterm infants. By 12 months of age, full-term infants, only, outperformed high-risk preterm infants. The authors concluded that all means were in the normal range; however, a significantly greater proportion of high-risk preterm infants obtained scores that fell one or two SDs below the normative mean. Using the same group assignments, Landry, Denson and Swank (1997), found that infants in the high-risk preterm group performed more poorly on the Mental Scale than both the low-risk preterm and full term groups at 6 months. This difference did not persist at 12 or 24 months. In a study by Gross and colleagues (1992), full term infants matched on gender, race, maternal age, education, and marital status outperformed preterm infants on the BSID Mental, but not Motor Scale, throughout the first two years of life.

Cooper and Sandler (1997), who grouped infants by birthweight, and Brady, Crowe and Deitz (1992), who grouped infants by size for gestational age, did not find any significant differences in group scores. Lastly, in a small study by Medoff-Cooper and Gennaro (1996), infants with abnormal sucking patterns obtained significantly lower PDI scores than infants with normal sucking patterns.

Other studies assigned infants to groups based on the degree of environmental or sociodemographic risk. Thompson, Oehler, Catlett, and Johndrow (1993) assigned infants into high distress and low distress maternal adjustment groups. Infants of mothers with high distress had significantly lower MDI and PDI scores.

Resnick, Armstrong, and Carter (1988) grouped infants into treatment and contrast groups. The treatment group received intervention using a preventative approach and the contrast group received traditional remedial intervention. There were no significant differences in MDI or PDI scores at 6 months of age between the groups; however, by 12 months of age the treatment group obtained significantly higher MDI scores, suggesting that treatment focused on reducing environmental risks may be more effective.

Studies by Youngblut, Loveland-Cherry, and Horan (1991;1993) grouped infants by degree of environmental risk based on maternal employment status. No significant differences were found in either study between the groups at any of the assessment times.

Although results of the studies reviewed were not consistent, there appears to be an overall trend of lower performance on the BSID for preterm infants of higher biologic or environmental risk. It is possible that preterm infants perform better cognitively than motoricly.

Related Studies Using the Revised Bayley Scales

A clinical validity study presented in the BSID-II manual included a sample of 57 preterm children with moderate medical risk. Mean MDI and PDI scores were based on corrected ages and were reported to be 88.6 (SD=15.7) and 83.5 (SD=21.6) respectively (Bayley, 1993). It was concluded that, in relation to the normative sample, children born prematurely performed more poorly. It was not reported whether these differences in scores were statistically or clinically significant.

As part of a study by D.J. Goldstein and colleagues (1995), the BSID-II was administered to 49 high-risk preterm infants. The infants' corrected ages were used to determine test starting point and scoring. Mean MDI score was 92.77 (SD=15.80) and mean PDI was 83.00 (SD=16.55). It was reported that 22% of the children obtained MDI scores which were classified as mildly or significantly delayed performance (as defined by the BSID-II classification guidelines), whereas 49% obtained PDI scores which fell within these ranges. Although there appeared to be a trend that preterm children performed better on the Mental compared to the Motor Scale, it was not examined whether these differences were statistically significant.

In a study by Macias and colleagues (1998), the BSID-II Mental Scale was used as a criterion measure to evaluate two infant screening tools. The BSID-II was administered to 78 infants born prematurely and/or with perinatal insult. Six of these infants were exposed to cocaine in utero. The infants were assessed between 6 and 24 months of age (corrected for infants born before 36 weeks gestation). Mean MDI score was 91.6 (SD=17.0). It was reported that 14.1% of the sample scored below 70 and 25.6% scored below 85. Therefore, approximately one-quarter of the premature infants

would have been classified as having a mild or significant delay in cognitive development.

As part of a study by Case-Smith and colleagues (1998), the BSID-II Mental and Motor Scales were administered to 45 preterm infants at 12-months adjusted age. Mean MDI was 97.74 (SD=15.5), and the mean PDI was 89.90 (SD=20.8). Although the PDI mean score was .67 standard deviations below the norm, motor performance was considered to be within normal limits. The authors concluded that motor development appeared to be more affected by premature birth than cognitive development.

Doig and colleagues (1999) administered the BSID-II Mental Scale to 38 high-risk infants. The sample included predominantly preterm infants with significant perinatal medical concerns, some in conjunction with prenatal cocaine exposure (16%). The mean age at the time of testing was 25.5 months (ranging from 15 to 40 months). Fifty-three percent of the sample were classified as performing within normal limits (MDI>85), 19% as mildly delayed (MDI between 70 and 85), and 28% as significantly delayed (MDI<70).

Due to the limited number of studies published that used the BSID-II with infants born prematurely and/or with significant perinatal medical concerns, it is difficult to draw solid conclusions. However, preterm infants appeared to perform better cognitively than motorically as measured by BSID-II scores.

Measuring Developmental Change In High-Risk Infants

Measuring developmental change not only requires an accurate measurement instrument, but also reasonable stability of development over time. However, the nature of development in infancy and childhood is considered to be unstable, variable and

changing (Aylward et al., 1987; Cole & Harris, 1992; Coryell et al., 1989; Darrah et al., 1998; Dunst & Rheingrover, 1981). As infants can go through rapid and slow changes in development at different time periods, instability in test scores may be a reflection of the differences in rate and timing of growth (Coryell et al., 1989). Therefore, sequential or serial testing of infant development have been recommended (Campbell et al., 1986; Darrah et al., 1998; Piper et al., 1991). When interpreting developmental change over time in high-risk infants, the nature of the subjects assessed must also be considered (Coryell et al., 1989).

Because norm-referenced tests have norms based on typically developing children, and as the performance of infants suspected of having a developmental delay may not experience typical changes in behavior, instability in test scores may be even greater for high-risk infants. Furthermore, high-risk populations may be more vulnerable to the influence of particular biologic and/or environmental factors (i.e. increased trait instability).

As there are only a limited number of studies published in which BSID-II data were collected on more than one occasion, studies using the BSID were reviewed. Caution should be used when evaluating changes in BSID test scores over time because results using the BSID-II will not necessarily demonstrate the same trends. Again the literature review will be divided into exposed and preterm infants.

Infants Prenatally Exposed to Drugs

Related Studies Using the Original Bayley Scales

Seven of the nineteen studies reviewed collected data using the BSID longitudinally (refer to Appendix A). Richardson and colleagues (1995) studied children

prenatally exposed to alcohol at 8 and 18 months of age. The mean MDI score at 8 months was more than one SD above the normative mean. At 18 months, the MDI mean dropped to approximately one-half SD above average. It was not reported whether this decrease in scores was statistically significant. Mean PDI scores remained more than one-half SD above average at both assessment points. In a study by Fried and Watkinson (1988), the BSID Mental Scale was administered to alcohol-exposed children at 12 and 24 months of age. There was no significant change in MDI scores over time. Chasnoff and colleagues (1992), found a higher incidence of delays (i.e. children with scores one or more SD below the normative mean) beginning at 6 months of age for children with prenatal exposure to alcohol and/or marijuana compared to a control group. However, trends in development over time were not described.

Three studies investigated developmental change in children prenatally exposed to opiates (Chasnoff et al., 1986; Rosen & Johnson, 1982; Wilson, 1989). Chasnoff and colleagues (1986), and Wilson (1989) reported a downward trend in scores over time for all groups, including the control, used in the studies. It was suggested that this downward trend was typical for children from lower socioeconomic populations, such as the children in their study samples. It was further suggested that the environment might have a greater influence on later development than prenatal drug exposure. In the study by Rosen and Johnson (1982), a downward trend in BSID test scores over time occurred for the exposed, but not the for the control group. The gap between mean scores for the two groups increased over time (i.e. at 12 and 18 months of age), especially for PDI scores.

In a study by Hurt and colleagues (1995), MDI and PDI scores decreased significantly over time for both the cocaine-exposed and the control group children. As

with the studies on prenatal opiate exposure, it was suggested that the downward trend in developmental performance could be reflective of a low SES environment. When the percentage of children scoring one or more SDs below the mean (i.e. delayed performance) was used to compare groups, Chasnoff and colleagues (1992) found that the cocaine-exposed group had a higher percentage of children with delayed performance.

Overall, the studies reviewed had mixed findings. If there was a trend in developmental change, it was in a downward progression. At times, this downward trend in developmental performance coexisted for children in the control groups. Therefore, it is possible that environmental risk factors may have a greater influence on later development than the biologic risk associated with prenatal drug exposure.

Related Study Using the Revised Bayley Scales

Only one study (Alessandri et al., 1998) collected data using the BSID-II (Mental Scale only) on more than one occasion. A significant decrease in MDI scores from 8 to 18 months was obtained by high cocaine-exposed, low cocaine-exposed, and non-exposed groups ($p<.001$). Similar to results obtained by studies using the BSID, the downward trend in cognitive functioning may be more reflective of the environment than of the effects from the initial prenatal drug exposure.

Infants Born Prematurely

Related Studies Using the Original Bayley Scales

Seven of the studies reviewed reported findings on change in test scores over time (refer to Appendix B). In a study by Thompson and colleagues (1994), MDI and PDI scores significantly decreased between 6 and 24 months of age. Resnick and colleagues (1988) found a significant decrease in PDI scores between 6 and 12 months of age. Also,

a significant drop in MDI scores was reported for infants in their contrast group (i.e. infants receiving typical intervention services), but not for infants in their experimental treatment group. Alternatively, in a study by Youngblut, Loveland-Cherry, and Horan (1993), there was an increase in MDI scores from 3 to 9 months for preterm infants of unemployed mothers.

Oehler, Thompson, Goldstein, Gustafson, and Brazy (1996) assessed preterm infants, classified as high- or low- biological risk, at 6, 15, and 24 months of age. Subjects were further classified by type of delay based on performance classifications: no delay (MDI and PDI scores $\geq$85 at all assessment points), continuous delay (MDI or PDI scores <85 at all points), and late delay (MDI or PDI scores <85 at 24 months only). The authors concluded that high risk preterm infants and infants classified with a continuous delay were less adept in fine and gross motor skills through 24 months of age (means not reported).

Three of the studies used full-term control groups and divided preterm infants into high and low-risk groups based on the number and severity of perinatal medical complications (Anderson et al., 1996; Landry et al., 1997; Wildin et al., 1995). Anderson and colleagues (1996) found that the high-risk preterm infants had poorer cognitive and psychomotor performance compared to full-term infants up to one year of age, whereas low-risk preterm infants performed more poorly only up to 6 months of age. Therefore, it was suggested that low-risk preterm infants may "catch up" to full term infants by one year of age. In a study by Landry and colleagues (1997), high-risk and low-risk preterm infants demonstrated lesser gains in cognitive development between 6 and 12 months compared to full term infants. Between 12 and 24 months of age, only high-risk preterm

infants did not obtain increases in cognitive performance comparable to the other two groups. Therefore, it was concluded that high-risk preterm infants demonstrated a deceleration in cognitive growth, especially between the first and second year of life. Lastly, Wildin and colleagues (1995) found a decline in MDI and PDI scores between 6 to 12 months of age for preterm as well as full term infant groups. It was suggested that this universal decline in scores could be due to a lower SES environment.

In summary, there appeared to be a downward trend in scores over time for preterm infants with more severe medical complications at birth. Infants with milder birth complications appeared to catch-up over time. Environmental factors appeared to influence later developmental progress.

Related Study Using the Revised Bayley Scales

In an abstract by Mattia and deRegnier (1998) the BSID-II Mental and Motor Scales were administered on two occasions to a sample of extremely premature children. However, changes in scores over time were not analyzed.

Biologic and Environmental Variables Thought to Affect Development

Theoretical Framework

Determining long-term developmental outcomes is challenging because of the difficulty in controlling for biologic variables, which can be compounded by environmental variables. Therefore, Zuckerman and Bresnahan (1991) suggested a multifactorial developmental model which includes both prenatal effects (i.e. effects on the central nervous system, which are viewed as creating biologic vulnerability) and postnatal influences (i.e. importance of social environment and caretaking quality). Environmental influences can shape outcome due to the newborn's capacity for

adaptation. Therefore, biologic perinatal factors are thought to exert their influence primarily in early infancy, whereas environmental post-natal factors become more predominant in subsequent development. For example, although infants prenatally exposed to alcohol and/or other drugs may have perinatal risk factors due to the possible teratogenic action of the exposure, their development may more importantly be influenced by environmental risk factors. Lifestyle risks associated with women actively using alcohol, and especially illicit drugs, during pregnancy make it a "difficulty, if not impossibility, of separating out teratogenic effects of prenatal exposure to a drug from the negative consequences of growing up in a drug-using environment" (Day, Richardson, & McGauhey, 1994, p.204). Other environmental factors such as socioeconomic status, race, education, marital status, polydrug use, maternal health and age, maternal nutrition and prenatal care, geographic location, and degree of violence in the home, could also obscure the substance effect (Chasnoff, 1991; Day & Richardson, 1991; Johnson et al., 1997; Lindenberg & Keith, 1993).

Alternatively, the development of preterm high-risk infants may be more vulnerable to initial biologic or medical risk factors such as respiratory distress syndrome, intraventricular hemorrhage, respiratory and metabolic acidosis (Goldstein, Thompson, Oehler, & Brazy, 1995; Lewis & Bendersky; 1989; Minde et al., 1989). A favourable environment could help these infants overcome their initial lag in development.

Overall, the multifactorial developmental model suggests that outcome is determined, and predicted, by the dynamic interaction of the child and its environment. Consideration of all the factors, and how each factor modifies and potentiates the others, was recommended (Zuckerman & Bresnahan, 1991).

Methodologies Used

In the studies reviewed, investigating BSID or BSID-II outcomes of exposed and preterm infants, a variety of biologic and environmental variables were considered. Common biologic risk variables included gender, gestational age (GA), birthweight (BW), head circumference, birth order, parity, infant's age at testing, amount of prenatal care, obstetrical conditions, neonatal medical and/or neurobiologic complications. Common environmental risk variables included race; socioeconomic level; marital status; living arrangement (e.g. foster care placement); maternal age, education, IQ, and mental health. Unfortunately, number, type and measurement of these variables differed greatly among the studies, making comparison of results difficult.

To determine the relationships between biologic and environmental variables and developmental outcomes, a few different methods and statistical analyses have been used in the studies reviewed. Several studies explored relationships between risk variables and test scores through correlational analyses (Cooper & Sandler, 1997; Gennaro & Stringer, 1991; Golden et al., 1982; Gusella & Fried, 1984; Lipkin & Altshuler, 1994; O'Connor et al., 1986; Ross et al., 1992). Correlations are a measure of the linear relationship between two variables and do not imply causation. In the interpretation of correlation coefficients (r), direction and magnitude of the relationship is considered. A positive r-value indicates a direct relationship between the two variables, so that an increase in one variable is related to an increase in the other. A negative r-value indicates an inverse relationship, where an increase in one variable is related to a decrease in the other variable. Strength of the association is assessed by the magnitude of r. First, statistical significance should be assessed (i.e. the strength of the relationship between the two variables is significantly

greater than zero). However, a statistically significant relationship does not necessarily mean a clinically significant one, especially when the sample size is large (Domholdt, 1993). Interpretation should involve examination of the coefficient of determination, or r-squared (r^2), which is an indicator of the percentage of shared variance between the two variables.

Many studies used regression analyses to investigate the effects of risk variables on developmental outcomes (Bendersky & Lewis, 1994; Brazy et al., 1991; 1993; Brooks-Gunn et al., 1993; Chasnoff et al., 1992; Cooper & Sandler, 1997; Feingold, 1994; Fried & Watkinson, 1988; R.F. Goldstein et al., 1995; Holzman, Paneth, Little, & Pinto-Martin, 1995; Howard et al., 1995; Jacobson et al., 1993; Korner et al., 1993; Mellins et al., 1994; O'Connor et al., 1993; Richardson et al., 1995; Rosen & Johnson, 1982; Seagull et al., 1996; Streissguth et al., 1980; Thompson et al., 1993;1994; Wildin et al., 1995; Youngblut et al., 1991;1993). In simple linear regression, one independent variable, or predictor, is used to predict the level of the dependent variable, whereas in multiple linear regression (MLR) more than one predictor variable is used. As with correlation coefficients, regression coefficients (R) can be positive or negative in direction. After establishing statistical significance of the magnitude of R-values, R-squared (R^2) values are examined to determine the percentage of variance in the dependent variable that can be predicted by the independent variable(s). In MLR, partial R^2 values indicate the percentage of unique variance in the dependent variable that can be predicted by a particular independent variable. R^2 values should also be reported for the regression equation as an indicator of the total percentage of variance in the dependent variable, which can be predicted by the combination of all the independent variables.

Adjusted R^2 values are useful in interpreting regression equations where the number of predictors is large in relation to the sample size (Howell, 1997).

Lastly, a few studies further investigated the effect of risk variables on development by dividing the original sample into groups based on degree of risk and, subsequently analyzing group differences (Hans, 1989; Mellins et al., 1994; O'Connor et al., 1993; Richardson et al., 1995; Thompson et al., 1994).

Related Studies

Studies that investigated the relationship between risk variables and Bayley scores were reviewed. In order to look at the relationship between specific variables and outcomes, the results of the studies were grouped by common risk variables used, with biologic variables present first, follow by environmental risk variables. It was felt that grouping the review by individual risk variables would ease the decision of which variables to include in this study's analyses. If correlations were used to investigate relationships in the studies, r^2-values are presented. If MLR analysis was used, partial R^2-values are reported. If analyses of group differences were used, group assignment was reported.

Typically-Developing Infants Using BSID-II

Two studies investigated relationships to BSID-II test scores using samples of typically-developing children (Levy-Shiff et al. 1998; Saudino et al, 1998). Levy-Shiff and colleagues (1998) tested 140 first born infants at 12 months of age. They found that maternal education was directly, and significantly ($p<.05$), predictive of MDI ($R = 0.23$) and PDI ($R = 0.23$) scores. Maternal education was also indirectly predictive of all three outcomes through its relationship with maternal efficacy ($R = 0.37$).

Saudino and colleagues (1998) administered the BSID-II Mental Scale to 102 children who were either twins or triplets. Average age of testing was 2.2 years (SD =.26). A significant correlation ($p<.001$) was found between age at time of testing and MDI score ($r =0.32$). The authors suggested that age standardization of the MDI might not be complete because the standardization of the BSID-II for ages 12 to 30 months is based on 3-month age spans, with scores for intermediate ages being derived from interpolation. Furthermore, because standard scores are provided for one-month intervals, they may not capture change that can take place within that time span.

Infants Prenatally Exposed to Drugs

Thirteen studies examined the relationship between biologic variables and scores obtained by infants prenatally exposed to drugs using the original Bayley Scales (refer to Appendix A). Depending upon the method of analysis, prenatal alcohol and/or drug exposure were found to relate inversely, and/or negatively predict BSID test scores in several studies (Gusella & Fried, 1984; Jacobson et al., 1993; O'Connor et al., 1986; Richardson et al., 1995; Rosen & Johnson, 1982; Streissguth et al., 1980). In some studies, alcohol and/or drug exposure were predictive or associated with MDI, but not PDI, scores (Fried & Watkinson, 1988; Mellins et al., 1994; Seagull et al., 1996). However reported magnitudes of the relationships, or the predictive strengths, tended to be weak (r^2 values from 0.04 to 0.41 and, partial R^2 values from 0.02 to 0.15). In other studies, no significant relationship was found between alcohol and/or drug exposure and BSID scores (Howard et al., 1995; O'Connor et al., 1993).

The relationship between gender and developmental outcome was investigated in six studies (Chasnoff et al., 1992; Johnson et al., 1997; Mellins et al., 1994; Richardson et

al., 1995; Rosen & Johnson, 1982; Seagull et al., 1996). Females scored lower than males on the Mental Scale at 12 months of age in studies by Seagull and colleagues and Richardson and colleagues, whereas Rosen and Johnson found the opposite. Unfortunately, magnitudes of correlation/regression coefficients were not provided in any of these studies, making interpretation difficult. No significant differences in test scores by gender were found in the other studies (Chasnoff et al., 1992; Johnson et al., 1997; Mellins et al., 1994).

Two studies (Gusella & Fried, 1984; Howard et al., 1995) found a very small direct relationship between birthweight and mental and motor scores (r^2 ranging from 0.04 to 0.10). This relationship was not significant in a study by Rosen and Johnson (1982).

Streissguth and colleagues (1980) reported that gestational age was the strongest predictor of MDI and PDI scores (partial R^2 value not reported) of the variables which entered into the equation (i.e. prenatal alcohol, nicotine and caffeine intake). Older gestational age led to higher test scores. However, in studies by Howard and colleagues (1995) and Rosen and Johnson (1982), gestational age did not relate significantly to test scores.

Jacobson and colleagues (1993) and, Richardson and colleagues (1995) found an inverse relationship between infant's age at testing and MDI and PDI scores (r^2 ranging from 0.04 to 0.07). Mellins and colleagues (1994) found this relationship to be statistically significant for MDI (r^2= 0.07), but not PDI scores.

Parity was a statistically significant predictor of 12-month MDI scores (partial R^2 =0.02) along with five other variables in a study by Fried and Watkinson (1988).

Increased parity was associated with 12-month PDI scores (r^2=0.02) in a study by Jacobson and colleagues (1993).

The relationships between environmental risk variables and BSID test scores were investigated in ten of the studies reviewed (refer to Appendix B). Quality of the home environment, as measured by the Home Observation for Measurement of the Environment, or HOME, Inventory (Caldwell & Bradley, 1979), was the strongest predictor of 24-month MDI scores (partial R^2= 0.08) when prenatal alcohol and nicotine intake, maternal education and calorie intake, and parity were the other predictors in a study by Fried and Watkinson (1988). HOME scores related directly to PDI and MDI scores (r^2 values ranging from 0.02 to 0.04) in a study by Jacobson and colleagues (1993). Howard and colleagues (1995) did not find a significant relationship between HOME and BSID scores.

The effect of race on developmental outcome was investigated in three studies (Billman et al., 1996; Mellins et al., 1994; Richardson et al., 1995). Richardson and colleagues found that race was one of 10 variables which entered into the analysis to predict 18-month MDI scores (partial R^2 value not reported). Being of white versus black race led to higher scores. Alternatively Billman and colleagues found black infants outperformed white infants on the Psychomotor Scale (significant group difference at $p<.05$ level). Mellins and colleagues found that BSID performance was not predicted by race.

Gusella and Fried (1984) found a weak, direct relationship between paternal education and MDI scores (r^2= 0.04). Seagull and colleagues (1996) did not find a significant association between maternal education and BSID outcomes. O'Connor and

colleagues (1993) also found no significant relationship between maternal intelligence and test score. A small inverse relationship was found between maternal age and MDI scores (r^2= 0.04) in a study by Jacobson and colleagues (1993).

PDI scores were inversely and weakly related to maternal depression (r^2= 0.02) in a study by Jacobson and colleagues (1993). Richardson and colleagues (1995) found that lower maternal depression predicted higher 8-month PDI scores (partial R^2 value not reported) in conjunction with age of infant, current infant weight, infant hospitalizations, gender, and current maternal work/school status.

In a study by Hans (1989), the combination of prenatal methadone exposure and lower SES led to lower test scores (significant group difference at $p<.05$). Seagull and colleagues (1996) did not find an association between maternal income and test scores.

Alcohol intake prior to pregnancy was not associated with lower test scores in a study by Gusella and Fried (1984), but was associated with lower MDI scores (r^2=0.35) in a study by O'Connor and colleagues (1986).

Other environmental risk variables were found to be associated with, or predictive of test scores in isolated instances (refer to Appendix A).

In summary, the effect of biologic risk variables on developmental outcomes as measured by the original Bayley Scales was mixed. When statistically significant relationships did exist, the shared, or accounted for, variance in BSID scores tended to be low. Study results on the effects of environmental risk variables were mixed. When relationships existed, their association, or predictive strength, tended to be weak. Therefore, the statistically significant relationships found may not represent clinically

significant relationships. Interpretation of the strength of predictor variables was not possible when neither r^2 nor partial R^2 values were reported.

Only one study has investigated relationships to developmental outcomes using the revised Bayley Scales (Alessandri et al., 1998). A regression analysis performed in the study by Alessandri and colleagues (1998) revealed that 8-month PDI scores were significantly predicted by group status (i.e. high cocaine exposure, low cocaine exposure, and no cocaine exposure), exposure to other toxic substances, and medical and environmental risk (R^2=0.17). Group status being the only significant independent predictor (partial R^2=0.11). Although this set of variables did not significantly predict 8-month MDI scores, environmental risk was an independent significant predictor (partial R^2=0.06). Eighteen-month MDI scores were predicted from the variables (R^2=0.14), with group status and environmental risk making significant independent contributions (partial R^2=0.05 and 0.04 respectively). Change in MDI scores was not predicted by this set of variables, although group status was a significant independent predictor (partial R^2=0.11). The authors concluded that environmental risk was related to cognitive functioning at both 8- and 18- months of age. However, because the environment appeared to have a lower impact on development for infants with less exposure to cocaine, it was suggested that cocaine-exposed infants should be viewed as a heterogeneous group. Infants with high levels of prenatal cocaine exposure may be more vulnerable to environmental risk factors and more likely to exhibit difficulties in cognitive functioning, especially at an older age.

Infants Born Prematurely

Fourteen studies examined the relationships between risk variables and BSID test scores (refer to Appendix B). Increased number and/or severity of perinatal medical complications was predictive of lower MDI and PDI scores in several studies (Bendersky & Lewis, 1994; Brazy et al., 1991; 1993; Korner et al., 1993; Landry et al., 1997; Thompson et al. 1994) but not in a study by Ross and colleagues (1992). When partial R^2 values were reported, they ranged from 0.13 to 0.32, and increased to a range of 0.23 to 0.39 when only infants with MDI or PDI scores below 85 were considered. Korner and colleagues (1993) found that their measure of neonatal medical complications was predictive of later PDI scores and of later MDI scores (partial R^2 not reported) for infants with lower birthweight. Neurodevelopmental risk examination was negatively related to MDI scores at term (r^2= 0.14) but not to later development in a study by Lipkin and Altshuler (1994). As the above-mentioned studies used different methods to determine degree of biologic risk, direct comparison of results is difficult.

Other studies used more specific indicators to represent biologic risk. Increased severity of IVH (through group assignment) led to lower test scores in studies by Bendersky and Lewis (1994), and Ross and colleagues (1992). In a study by R.F. Goldstein and colleagues (1995), metabolic acidosis and hypotension variables combined predicted lower MDI scores (R^2=0.15) and lower PDI scores (R^2=0.20).

In a study by Cooper and Sandler (1997), porencephaly was the strongest negative predictor of total Bayley scores (R-values not reported) of all variables that entered into the analysis (i.e. inadequate intrauterine growth, more days on supplemental oxygen, and

lower maternal education). In a study by Brady and colleagues (1991), size for gestational age was not associated with 6-, 15- or 24-month test scores.

Gender was the second strongest predictor (next to number and severity of medical complications) of test scores (partial R^2 ranging from 0.04 to 0.10) in studies by Brazy and colleagues (1993) and Thompson and colleagues (1994). Females tended to have higher MDI and PDI scores than males. However, in other studies, gender did not affect test scores (Gross et al., 1992; Ross et al., 1992).

Birthweight was not associated with test scores in the studies by Brazy and colleagues (1991; 1993) and by Feingold (1994). Gestational age was a significant predictor of 6-month PDI scores (partial R^2= 0.02) along with cumulative medical risk measure and gender in a study by Brazy and colleagues (1993). However this relationship was not significant in two other studies (Brazy et al., 1991; Gross et al., 1992).

Although study results were inconsistent, measurement tools that measured the number and severity of commonly occurring perinatal medical complications in a cumulative fashion were more likely to be related to developmental outcome, than did individual reflectors of medical status such as gestational age or birthweight. Furthermore, there may be a trend that preterm female infants have more favorable outcomes compared to males.

The relationships between environmental variables and BSID test scores were investigated in eleven of the studies reviewed (refer to Appendix B). In a study by Bendersky and Lewis (1994), family risk, and the interaction between family/environmental risk by IVH severity predicted MDI, but not PDI scores (partial R-

values not reported). Higher HOME Inventory scores were associated with higher MDI scores (r^2= 0.29) in a study by Feingold (1994).

Race (white vs. black) predicted higher 15- and 24- month MDI scores (partial R^2 =0.03 and 0.07 respectively) in a studies by Brazy and colleagues (1993) and Thompson and colleagues (1994). Similarly in a study by Korner and colleagues (1993), black or Hispanic race led to lower 24-month MDI scores (partial R-values not reported) compared to white race. However, race did not relate to test scores in a study by Ross and colleagues (1992).

In a study by Brazy and colleagues (1993), higher maternal educational level predicted higher 24-month MDI (along with cumulative medical risk and race) and higher 24- month PDI scores (along with cumulative medical risk and gender). Partial R^2 value were 0.17 and 0.04 respectively. Cooper and Sandler (1997) found maternal education to be a predictor of both MDI and PDI scores (partial R-values not reported). Thompson and colleagues (1994) found that higher maternal education predicted higher 24- month PDI scores only (partial R^2=0.04), whereas Korner and colleagues (1993) found this predictive relationship for 24-month MDI scores only (partial R^2 not reported). No significant relationship between maternal education and BSID scores was found in other studies (Feingold, 1994; Gross et al., 1992).

Thompson and colleagues (1994) found that higher SES led to higher 24-month MDI scores (partial R^2= 0.16), whereas Ross and colleagues (1992) found no significant relationships between social class and SES, and test scores.

Maternal stress was found to be a predictor of MDI scores at 6, 15 and, 24 months (partial R^2 ranged from 0.06 to 0.10) in a study by Thompson and colleagues (1994).

Maternal depressive symptoms were not significantly related to test scores in studies by Feingold (1994) and Gennaro and Stringer (1991).

Although, Youngblut and colleagues (1991) found a positive relationship between maternal hours of employment and PDI scores (significant group differences at $p<.05$), this was not replicated in a later study of theirs (Youngblut et al., 1993).

In summary, the study results using the original Bayley Scales with infants born prematurely were mixed and inconsistent. In general, biologic variables were more predictive of later motor outcomes and environmental variables were more predictive of later cognitive outcomes.

Two studies used the revised Bayley Scales as an outcome measure with infants born prematurely (Costarides & Shulman, 1998; Mattia & deRegnier, 1998). Costarides and Shulman (1998) studied the relationship between the BSID-II Mental Scale and a norm-referenced language measure for a sample of 90 infants at risk for a developmental delay due to prematurity, low birth weight, or physical condition. There was a significant ($p<0.05$) correlation between the language test scores and the 12 and 24 month MDI scores ($r =0.30$ and $r =0.31$ respectively). Because the Bayley Mental Scale includes language items, these results are not surprising.

The BSID-II Mental and Motor Scales were administered to 96 extremely premature infants (gestational age $\leq$ 30 weeks) at 1, 2 and 3 years of age (Mattia & deRegnier, 1998). Infants with the highest degree of physiologic instability had significantly lower MDI scores at 1 year of age and lower PDI scores at 1, 2 and 3 years of age (means not reported). Multiple regression analyses revealed that a higher degree of physiologic instability, more severe intraventricular hemorrhage, and lower gestational

age were associated with lower 1-year MDI scores (R-values not reported). A higher degree of physiologic instability and more severe intraventricular hemorrhage were associated with lower 1-year PDI scores. By 2 years of age, only degree of physiologic instability was associated with MDI and PDI scores.

Although a limited number of studies has been published, it appears that the degree of perinatal biologic risk may affect earlier cognitive and motor performance but have long-term effects on motor performance only. There are no published studies available that investigate relationships between environmental risk variables and preterm infants' performance on the BSID-II.

Summary

Although there were only a few published studies that have used the BSID-II with infants prenatally exposed to alcohol and/or other drugs, the following trends may exist and warrant further investigation: (1) better motor than cognitive performance (Bayley, 1993), (2) development of cognitive skills may be delayed, but not identified by the BSID-II until later in life (at 18 months-of-age), (3) cognitive scores may decrease over time, especially for infants with higher biologic and environmental risk, and (4) environmental risk may be related to cognitive performance in infants as early as 8 months of age. In reviewing studies using the original Bayley Scales with infants prenatally exposed to drugs, the results were inconsistent. If a difference in performance did exist between exposed infants and control groups, the control groups tended to outperform the exposed groups. If a significant change in scores occurred over time, it was in a downward progression (i.e. although raw scores may have increased, index scores decreased). At times this downward trend in developmental performance coexisted

for children in the control groups. Therefore, it is possible that environmental factors may have a greater influence on later development than prenatal drug exposure.

Review of the limited studies that used the BSID-II to investigate development of infants born prematurely and/or with significant perinatal medical concerns revealed the following trends: (1) these infants may have delays in cognitive and motor development with delays tending to be more severe in motor development, and (2) the degree of perinatal biologic risk may affect earlier cognitive and motor performance but have greater long-term effects on motor performance only. In a review of studies using the original Bayley Scales with infants born prematurely, results were inconsistent. A trend of lower performance for preterm infants of higher biologic and environmental risk may exist. There appeared to be a downward trend in scores over time for preterm infants with more severe medical complications at birth. Infants with milder birth complications may "catch-up" to typically developing children over time. Biologic risk variables appeared to be more predictive of later motor outcomes and environmental risk variables appeared to influence later cognitive outcomes.

CHAPTER 3

Methods

The following section describes the methodology used in this study to (1) investigate the ability of the BSID-II to identify infants at risk of a developmental delay, (2) describe changes in BSID-II test scores over time, and (3) explore potential relationships between biologic and environmental risk variables and BSID-II outcomes.

Research Questions

Based on the purposes of the study and the review of the literature, the following research questions were developed:

Question 1: Identification of Developmental Delay

(a) Did the sample of high-risk infants score significantly lower on the Bayley-II Motor and Mental Scales compared to the norms?

(b) What percentage of the high-risk infants was classified, using the BSID-II recommended performance classifications, as having a mild or significant developmental delay at each testing occasion?

Question 2: Change in Performance Over Time

(a) Did Bayley-II MDI and PDI scores change significantly from the first to the second testing occasion for each group (i.e. infants prenatally exposed to alcohol and/or other drugs and infants born prematurely and/or with significant perinatal medical concerns)?

(b) How did performance classifications change from the first to the second testing occasion for each group?

Question 3: Exploration of Relationships Between Risk Variables and Performance

(a) Do any relationships exist between the infants' scores and biologic and/or environmental risk variables?

Participants

BSID-II test scores were collected as part of a larger study funded by the British Columbia Health Research Foundation Grant # 146 (95-1) from 1993 to 1995 (Harris, 1995). A convenience sample of 53 infants was recruited from two sites in Vancouver, British Columbia, each with a developmental follow-up programme: Sunny Hill Health Centre for Children (SHHCC) and British Columbia Children's Hospital (BCCH). At SHHCC, BSID-II scores were collected on 34 infants who had been prenatally exposed to alcohol and/or other drugs, 14 of whom had also been born prematurely (i.e. gestational age of < 37 weeks). All but one infant were admitted from their birth hospital to SHHCC for treatment of Neonatal Abstinence Syndrome (i.e. withdrawal symptoms). It was requested that the other infant be tested prior to adoption due to concerns of prenatal drug exposure. At BCCH, BSID-II scores were collected on 19 infants who were born prematurely and/or were considered to have significant perinatal medical factors (e.g. intraventricular hemorrhage, respiratory distress syndrome, periventricular leukomalacia, bronchopulmonary dysplasia, bradycardia, asphyxia, seizures, apnea, hydro/microcephaly, patent ductus arteriosus). Of these 19 infants, 13 were born prematurely. The remaining 6 infants had severe meconium aspiration necessitating extracorporeal membrane oxygenation (ECMO) therapy.

The BCCH group was born at a significantly younger GA and lower BW than the SHHCC group (refer to Table 1). There was a significantly greater proportion of male

infants in the BCCH group compared to the SHHCC group ($p<.05$). Although in general it is suspected that slightly more preterm males have significant perinatal medical concerns compared to females (Corocoran, Patterson, Thomas, & Halliday, 1993; Todd, Jana, & John, 1997), this difference is not expected to be as large as it is in the BCCH group (i.e. 74%). However, there were no significant differences in index score means between males and females in this study. Attrition occurred with four infants (7.5 %) at SHHCC and none at BCCH. Reasons for attrition were not documented in the medical charts.

The BSID-II was administered by four different examiners, two at each of the sites, following administration of the Harris Infant Neuromotor Test (Harris, 1993). The examiners were three physical therapists and one occupational therapist who were employed by the participating sites. Corrected ages were used for assessment and scoring of infants born at, or less than, 37 weeks' gestation. All examiners had previous training and experience in administering and scoring the first edition of the Bayley Scales. As well, training and inter-rater reliability evaluations were conducted using the BSID-II. Inter-rater reliability for the raw scores on the BSID-II was obtained using six infants without risk concerns, ranging form 3 ½ to 18 months. Intraclass correlation coefficients (ICC) were .993 for the Mental Scale and .995 for the Motor Scale (Harris, 1995).

Table 1

Participant Characteristics

Variables	Total Sample	SHHCC Group	BCCH Group
Infant Variables			
Gestational Age (weeks)			
Mean	35.67	37.67	32.21 ***
SD	4.92	2.43	6.17
Birthweight (grams)			
Mean	2493.87	2753.38	2029.47 ***
SD	999.88	692.87	1285.78
Gender [% (n)]			
Male	64.2 (34)	58.5 (20)	73.7 (14) *
Female	35.8 (19)	41.2 (14)	26.3 (5)
Maternal Variables			
Age (years)			
Mean	28.70	27.39	30.84
SD	5.99	6.14	5.21
Ethnicity [% (n)]			
Caucasian	45.3 (24)	47.1 (16)	42.1 (8)
Non-Caucasian	45.3 (24)	38.2 (13)	57.9 (11)
Missing	9.4 (5)	14.7 (5)	0

Note. SHHCC = Sunny Hill Health Centre for Children, or prenatal drug-exposure group; BCCH = British Columbia's Children's Hospital, or preterm and/or with perinatal medial concerns group.

* $p<.05$. *** $p<.001$ significant difference between SHHCC and BCCH groups

Study Design and Procedure

This retrospective study received ethical review approval from the Clinical Research Ethics Board at the University of British Columbia, as well as by the research review committee for BCCH and SHHCC. The participants' medical charts were reviewed to verify the accuracy of the initial data collection by Harris (1995) and to collect sociodemographic information, including biologic and environmental variables.

Data Analyses

All data analyses were performed using the Statistical Software Package for the Social Sciences 9.0 (SPSS). Data distribution and extreme scores (i.e. outliers) were evaluated prior to statistical analyses. In order to include as much of the sample and representation as possible, extrapolated scores were derived for MDI and PDI scores that fell below 50 using information by Robinson and Mervis (1996). Robinson and Mervis provided a table with extrapolated scores for values 30 to 49. The formulae provided by Robinson and Mervis for values below 30 led to negative scores in some instances. Theoretically, it is not possible to have negative development. Because the extremely low scores obtained by a few infants were considered to reflect true performance and not errors, a minimum value of 15 was assigned to these. An index score of 15 is six SDs below the mean, and therefore, represents a very extreme case.

Question 1(a): Identification of Developmental Delays Through Comparison of Scores to Normative Data

BSID-II MDI and PDI means obtained from the total sample at each testing occasion were compared to the normative mean. Analyses using MDI and PDI scores, including extrapolated scores (MDIex and PDIex respectively) at each testing occasion,

were performed separately. A statistically significant difference in means was determined using eight one-sample t-tests. Because multiple t-tests were conducted, the alpha level was set at .01. Significance testing was one-tailed as it was hypothesized the high-risk infants would score lower than typically-developing infants. The null hypothesis (H_0) tested was that the population mean of high-risk infants was equal to or greater than the normative mean.

Only one of the studies reviewed compared sample means to normative means (Chasnoff et al., 1992). One-half a SD was chosen as an indication of a clinically significant difference. An effect size of one-half SD (i.e. a drop of 7.5 points in MDI or PDI means) was chosen to represent a clinically significant difference this study as well. This effect size was chosen for comparing group means because it is larger than the size of the standard error of measurement, but not as stringent as a difference of one standard deviation used for interpreting a delay for individual test scores. A group mean of one-half a SD below the norm would likely include several individuals who scored one or more SDs below the normative mean.

Power, or the ability to detect a difference when a true difference exists, of the t-tests was calculated using GPOWER (Faul & Erdfelder,1992). With a sample size of 53, effect size of .5 SD, and alpha level of .01, the power of the unidirectional t-tests was calculated to be 88. Power was considered to be high.

Although it would have been of interest to compare the means obtained from each subgroup to the normative mean, this was not done in order to reduce the number of tests performed. One of the main purposes of the BSID-II is to help identify infants at risk of a developmental delay, regardless of the cause of the delay (Bayley, 1993).

Question 1(b): Identification of Developmental Delays Through the Use of Performance Classifications

The incidence of mildly and severely delayed cognitive and motor performance was determined for each testing occasion. To increase reliability of the classifications and to reflect clinical practice, each infant's MDI and PDI scores were converted to a 95% confidence interval (CI) using the Standard Error of Measurement (SEM) provided in the BSID-II manual. The resultant test score range was then converted to the BSID-II recommended performance classifications. Percentage of mildly and significantly delayed performance classifications at each testing occasion was calculated for the entire sample.

Question 2(a): Description of Change in Test Scores Over Time Through Comparison of Means

Four Group (2) by Time (2) Analyses of Variance (ANOVAs) were performed, one for each of the following dependent variables: Mental Developmental Index scores (MDI), Mental Developmental Index scores including extrapolated scores (MDIex), Psychomotor Developmental Index scores (PDI), and Psychomotor Developmental Index scores including extrapolated scores (PDIex). The ANOVAs were run using the general linear model (GLM). This procedure is more powerful than Factorial models (SPSS Base 9.0 Applications Guide, 1999). Furthermore, the GLM approach was reported to work well with cases of unequal sample sizes (Howell, 1997) like in this study. Four ANOVAs versus two multiple analyses of variance (MANOVA) were chosen because the BSID-II Mental and Motor Scales were designed to measure different constructs. Groups were infants prenatally exposed to alcohol and/or other drugs, and infants born prematurely

and/or with significant perinatal medical factors. Time was the first (i.e. before 12 months of age) and the second (i.e. around 18 months of age) testing occasion. Due to the retrospective design of this study there was no control for confounding variables, such as type and amount of intervention.

None of the articles reviewed reported an effect size for change of an individual performance over time. It is difficult to select what would be a meaningful change in an index score because the item sets would differ from one testing occasion to the next. Also, the items within an item set would differ in the amount of meaningfulness to the child and caregivers. Therefore, a change in group mean of one-half a standard deviation was considered clinically significant. A change in group mean of such an amount would include several infants whose performance on the BSID-II improved at a rate greater than what is expected through normal growth and development.

A priori power estimations were calculated using formulae provided by Park and Schutz (1999). Assuming a sample size of 53, mean correlations among test scores from time 1 to time 2 of .50, effect size of .50, and alpha level of .01, power was estimated to be .75 for the time main effect, .62 for the group main effect, and .33 for the interaction effect. Power was considered to be moderately high for the main effects but low for the interaction effect. Because this was a retrospective study, sample size could not be increased as a way to increase the power. If alpha level was set at .05 instead of .01, power estimations were calculated to be .94 for the time main effect, .81 for the group main effect, and .59 for the interaction effect. Due to the large number of statistical tests performed in this study, it was chosen to keep alpha level at .01. However, any

differences that would have reached statistical significance at an alpha level of .05 will be reported as possible directions for future research.

Question 2(b): Description of Change in Test Scores Over Time Through Changes in Performance Classifications

Incidence of change in performance classifications was determined for each group. Ninety-five percent confidence intervals (CI) were used to determine test score ranges and corresponding performance classifications. Percentages of no change, an increase, or a decrease in cognitive and motor performance classifications were calculated.

Question 3: Exploration of Relationships Between Risk Variables and Performance

Lastly, data analyses were performed to explore relationships between biologic and environmental risk variables and BSID-II outcomes. Relationships were explored through regression analyses. Risk variables used in the analyses were dependent on the availability of information in the infants' medical charts. Biologic risk variables used in the analyses were gestational age, birthweight, and gender. Because there was little documentation in the charts on environmental risk variables, especially for the drug/alcohol exposed group, only maternal age and maternal ethnicity were used.

Group status (infants exposed to alcohol and/or other drugs, and infants born prematurely and/or with significant perinatal medical concerns) was another risk variable used in the analyses. The nature of the two groups in the study was indicative of different risks. The infants admitted to BCCH were considered to be at higher biological risk than the infants admitted to SHHCC because all infants requiring intensive, acute pediatric medical interventions born in the province of British Columbia received care there.

Alternately, according to the literature reviewed, the subgroup of infants prenatally exposed to alcohol and/or other drugs are believed to be at higher environmental risk. However, because twenty-three (76%) of the drug/alcohol exposed infants were discharged to foster families, environmental risk was not necessarily greater for this group.

Because research results regarding the influence of certain biologic and environmental factors on development has been inconsistent and because theoretical perspectives are varied, stepwise multiple linear regression analyses were used. Dependent variables included MDI, MDIex, PDI, and PDIex scores at each testing occasion. Change in MDI, or ΔMDI, ΔMDIex, ΔPDI, and ΔPDIex were also used as dependent variables. Correlational analyses were performed between gender, birthweight, gestational age, maternal ethnicity, maternal age, group status, and BSID-II index scores at each testing occasion (MDI 1, MDI 2, PDI 1, PDI 2, MDIex 1, MDIex 2, PDIex 1, PDIex 2) and change in index scores from time 1 to time 2 (MDI, or ΔMDI, ΔMDIex, ΔPDI, ΔPDIex). Only the risk variables that correlated significantly ($p<.05$) and most highly with test scores were used as predictor variables in MLR. Due to the small sample size, a maximum of five independent variables was entered into the equation. The probability of F to enter a variable into the equation was set at $p\leq 0.10$, while the probability of F to remove a variable from the equation was set at $p\leq 0.15$. These somewhat lenient criteria were used because the investigations were of an exploratory nature. Power of MLR was calculated a priori using GPOWER (Faul & Erdfelder, 1992). With a moderate effect size ($f^2 = .15$), alpha level of .10, and sample size of 53, power

calculations ranged from .79 if two predictors entered into the equation to .64 if five predictors entered into the equation.

CHAPTER 4

Results

In this chapter, the results of each of the research questions will be reported. For ease in reading, infants prenatally exposed to alcohol and/or other drugs will be referred to as "exposed" infants and, infants born prematurely and/or with significant perinatal medical concerns will be referred to as "preterm" infants.

Question 1: Identification of Developmental Delays in High-Risk Sample

a) Comparison of Scores to Normative Data

Distributions of index scores were examined prior to administering statistical tests. T-tests have the assumption that the population is normally distributed (Howell, 1997). The distributions of MDIex and PDIex (extrapolated) scores were negatively skewed at each of the two testing occasions (refer to Table 2). However, it is important to recognize that large samples of data are needed before a good idea about the shape of the distribution can be made (Howell, 1997). With small sample sizes, marked skewness and kurtosis can be expected (Howell, 1997). Fortunately, the t-statistic is considered robust in regard to the normality assumption (Glass & Hopkins, 1984; Glenburg, 1988). Glass and Hopkins (1984) reported that "the violation of the assumption of normality has almost no practical consequences in using the t-test" (p. 237). Furthermore, they reported that the power of the t-test is virtually unaffected by marked non-normality.

Table 2

Descriptive Statistics for BSID-II Index Scores (Including Extrapolated Scores)

Descriptive	MDIex 1	PDIex 1	MDIex 2	PDIex 2
Number of scores (N)	53	53	47	49
Mean	85.11	79.79	75.55	73.94
Median	88	83	81	79
Standard Deviation	14.26	20.97	22.62	26.12
Skewness	-2.952	-1.251	-1.285	-1.114
Standard Error of Skewness	.327	.327	.347	.340
Kurtosis	11.418	2.079	1.913	.710
Standard Error of Kurtosis	.644	.644	.681	.668
Range	15-101	15-111	15-115	15-115

Note: BSID-II mean = 100 and standard deviation = 15

MDIex 1 = Mental Developmental Index scores including scores< 50 at time 1; MDIex 2 = Mental Developmental Index scores including scores< 50 at time 2; PDIex 1 = Psychomotor Developmental Index scores including scores< 50 at time 1; PDIex 2 = Psychomotor Developmental Index scores including scores< 50 at time 2.

When only MDI and PDI scores above 50, or non-extrapolated scores, were investigated, the distributions were less skewed (refer to Table 3). Clinically, index scores below 50 would simply be reported as such, indicating a very significant delay in development. For research purposes, it is better to include as much of the sample as

possible in the analyses. All statistical analyses were conducted using both extrapolated and non-extrapolated index scores as dependent variables in order to compare the results.

Table 3

Descriptive Statistics for BSID-II (Non-Extrapolated) Index Scores

Descriptive	MDI 1	PDI 1	MDI 2	PDI 2
Number of scores (N)	51	49	42	42
Mean	87.29	84.10	81.79	82.90
Median	88	85	83	79
Standard Deviation	8.63	14.59	13.47	14.35
Skewness	-1.267	-.226	.227	.121
Standard Error of Skewness	.333	.340	.365	.365
Kurtosis	1.633	-.258	-.030	-.452
Standard Error of Kurtosis	.656	.668	.717	.717
Range	62-101	53-111	59-115	52-115

Note: BSID-II mean = 100 and standard deviation = 15

MDI 1 = Mental Developmental Index scores $\geq$ 50 at time 1; MDI 2 = Mental Developmental Index scores $\geq$ 50 at time 2; PDI 1 = Psychomotor Developmental Index scores $\geq$ 50 at time 1; PDI 2 = Psychomotor Developmental Index scores $\geq$ 50 at time 2.

It was hypothesized that infants at risk for developmental delays (due to prenatal exposure to drugs or to premature birth and/or perinatal medical concerns) would obtain

lower MDI and PDI scores compared to the normative population. The null hypotheses were that high-risk infants obtained scores that were equal to, or greater than, the normative mean, H_0: $\mu \geq 100$ (where μ represents mental and psychomotor means at each testing occasion). One–tailed, one-sample t-tests with alpha level at .01 were used to compare MDIex, MDI, PDI, and PDIex scores at time one and two, to the BSID-II normative means. Using extrapolated scores, all null hypotheses were rejected with $t_{MDIex\ 1}(52) = -7.600$, $t_{PDI\ ex\ 1}(52) = -7.016$, $t_{MDIex\ 2}(46) = -7.408$ and, $t_{PDIex\ 2}(48) = -6.985$ at $p<.001$. There is strong evidence to suggest that infants prenatally exposed to alcohol and/or other drugs (exposed infants) and infants born prematurely and/or with significant perinatal medical concerns (preterm infants) obtained lower scores ($\overline{X}_{MDIex\ 1} = 85.11$, $\overline{X}_{PDIex\ 1} = 79.79$, $\overline{X}_{MDIex\ 2} = 75.55$ and, $\overline{X}_{PDIex\ 2} = 73.94$) on the BSID-II Mental and Motor Scales compared to the normative data. All sample means were more than one-half SD below the normative mean ($\overline{X}$ s < 92.5) and, therefore were considered to be clinically significantly different.

Similarly, when non-extrapolated index scores were analyzed, all null hypotheses were rejected with $t_{MDI\ 1}(50) = -10.510$, $t_{PDI\ 1}(48) = -7.626$, $t_{MDI\ 2}(41) = -8.763$ and $t_{PDI\ 2}(41) = -7.719$ at $p<.001$. As well, all the means ($\overline{X}_{MDI\ 1} = 87.29$, $\overline{X}_{PDI\ 1} = 84.10$, $\overline{X}_{MDI\ 2} = 81.79$ and, $\overline{X}_{PDI\ 2} = 82.90$) were considered to be clinically significantly different from the normative mean at both test times. Regardless of whether or not extrapolated scores were used, the BSID-II was able to identify exposed and preterm infants as being at risk for developmental delays.

b) Identification Through Use of Performance Classifications.

Use of extrapolated scores did not affect the BSID-II recommended performance classifications because all index scores of 69, or below, would be classified as significantly delayed performance. Twenty-eight percent of the sample was classified as having a mild or a significant delay in cognitive development at the first testing time (refer to Table 4). The percentage increased to 60% at time two. The majority of the sample was classified as having a mild or significant delay in motor performance at time one (55%) and time two (67%).

Table 4

Frequency of Performance Classifications

	Time 1				Time 2			
Performance	Mental		Motor		Mental		Motor	
Classification	n	%	n	%	n	%	n	%
Significantly delayed	6	11.3	14	26.4	15	31.9	16	32.7
Mildly delayed	9	17.0	15	28.3	13	27.7	17	34.7
Within normal limits	38	71.7	24	45.3	18	38.3	15	30.6
Accelerated	0	0	0	0	1	2.1	1	2.0
Total	53	100	53	100	47	100	49	100

When 95% confidence intervals (CIs) were used to interpret each individual's scores, often the resulting test score range would overlap into two performance classifications (refer to Table 5). The highest percentage of score ranges fell into the mildly delayed to normal range of performance. At under 12 months of age, only 13% of the sample was classified with a significant, or a significant to mild delay, in cognitive performance. However, if the mild delay to normal performance classification was included (i.e. mild delay to normal classification would translate to being at risk for a developmental delay), the resultant overall percentage rises to 72%.

Table 5

Frequency of Performance Classifications Obtained Using 95% Confidence Intervals

	Time 1				Time 2			
Performance	Mental		Motor		Mental		Motor	
Classification	n	%	n	%	n	%	n	%
Significant delay	1	1.9	4	7.5	5	10.6	8	16.3
Significant to mild delay	6	11.3	12	22.6	12	25.5	10	20.5
Mild delay to normal	31	58.5	25	47.2	23	48.9	19	38.8
Normal	15	28.3	9	17.0	6	12.8	10	20.5
Normal to accelerated	0	0	3	5.7	1	2.1	2	4.1
Total	53	100	53	100	47	100	49	100

At around 18 months of age, 37% of the sample was classified as having a significant, or a significant to mild delay, in cognitive performance. If the mild delay to normal performance classification is included, the percentage increased to 85%.

The percentage of the sample classified with a significant, or a significant to mild delay, on the Motor Scale, was about 30% at the first testing occasion and 37% at the second. Inclusion of the mild delay to normal performance classification raises the percentages to 77% and 76% respectively. Therefore, if infants classified with mild to normal performance are considered to be at risk for a development delay, along with the infants classified with a significant delay, and significant to mild delay, more of the sample would have been identified as having a possible delay in development using the 95% CIs. However, the chance of falsely identifying infants at risk for a developmental delay when they are developing normally, is increased. Because it would be more detrimental to withhold health care services to those individuals who needed it than to provide services to those who do not need it, use of CIs could be useful in making decisions regarding service delivery.

Question 2: Change in Scores For Each Group

(a) Change in Developmental Index Scores

It was hypothesized that (1) the exposed infants' scores (especially MDI scores) would decrease overtime, and, (2) the preterm infants PDI scores would increase over time. The null hypotheses were that the rate of developmental change was equal for exposed and preterm infants, H_0: $\mu_{SHHCC\,2} - \mu_{SHHCC\,1} = \mu_{BCCH\,2} - \mu_{BCCH\,1}$ (where $\mu_{SHHCC\,2} - \mu_{SHHCC\,1}$ represents the change in exposed infants' mental and psychomotor score

means and $\mu_{BCCH\ 2} - \mu_{BCCH\ 1}$ represents change in preterm infants' means). Furthermore, it was of interest to see if, overall, the scores differed between exposed and preterm infants, and if the high-risk infants, as a whole, had an increase or decrease in performance. Null hypotheses were H_0: $\mu_{SHHCC} = \mu_{BCCH}$ (where μ_{SHHCC} represents exposed infants' mean scores and μ_{BCCH} represents preterm infants' mean scores averaged over time) and H_0: $\mu_{time1} = \mu_{time\,2}$ (where μ_{time1} represents high-risk infants' mean scores at under 12 months of age and $\mu_{time\,2}$ represents high risk infants' mean scores at around 18 months of age).

Analysis of the distributions of the dependent variables divided into groups (i.e. infants prenatally exposed to drugs and/or alcohol, or SHHCC group, and infants born prematurely and/or with significant perinatal medical concerns, or BCCH group) revealed that the MDIex 1 distributions for both groups, and the MDIex 2 and PDIex 2 distributions for the SHHCC subgroup were negatively skewed (refer to Appendix C). Glass and Hopkins (1984) discussed that skewed populations have very little effect on either the level of significance or the power of an ANOVA.

Levene's test of Equality of Error Variance reached significance for PDIex 1, MDIex 2, and PDIex 2 ($p<.05$), and therefore, there is insufficient evidence to conclude that the variances between the two subgroups are equal. Glass and Hopkins (1984) reported that the level of significance might be higher than the actual level when the subgroup with the smaller number is also the subgroup with greater variability in scores, which is the case with these data.

When non-extrapolated scores were analyzed, the distributions for MDI 1 for both SHHCC and BCCH groups, were only slightly negatively skewed (refer to Appendix D). The Levene's test of Equality in Error Variance did not reach significance ($p>.05$) for any of the dependent variables (i.e. MDI 1, PDI 1, MDI 2 and PDI 2). Therefore, there is strong evidence to suggest that the variances were equal between the two groups for both scales and at both testing times. The use of non-extrapolated scores in the ANOVAs could result in more accurate significance levels, and therefore, more accurate interpretations. However, using extrapolated scores allowed for the inclusion of the more significantly delayed subjects' scores. Analyses were performed using both extrapolated and non-extrapolated scores and the results were compared. Alpha level was set at .01 instead of .05 to reflect the number of tests performed.

When MDIex scores were analyzed, there was a significant time main effect ($\overline{X}_{\text{MDIex 1}} = 85.18$ and $\overline{X}_{\text{MDIex 2}} = 74.94$; $F_{1,45}=12.036$, $p<.01$) providing strong evidence that scores were higher before one year of age compared to scores around 18 months of age for exposed and preterm infants (refer to Table 6 and 7). Group and, group by time interaction, effects were non-significant. Therefore, there is insufficient evidence to support differences in mental scores, or rate of change in mental scores, between exposed infants and preterm infants.

Table 6

Summary of Mental Developmental Index Means (Including Extrapolated Scores) for Time By Group ANOVA

		Time		
		<12 months $\overline{X}$ (SD)	18 months $\overline{X}$ (SD)	Marginal Means
Group	SHHCC	85.83 (11.53)	77.17 (14.63)	81.50
	BCCH	84.53 (18.82)	72.71 (32.71)	78.62
Marginal Means		85.18	74.94	

Table 7

Summary of ANOVA Results Using Mental Developmental Index Scores (Including Extrapolated Scores)

Source	SS	df	MS	F	p
		Between subjects			
Group	180.300	1	180.33	.333	.567
Error	24331.529	45	540.701		
		Within subjects			
Time	2277.899	1	2277.899	12.036	.001**
Time by Group	54.070	1	54.070	.286	.596
Error	8516.569	45	189.257		

**$p < .01$

When MDI scores were analyzed, there were no significant main or interaction effects (refer to Table 8 and 9). As such, there is insufficient evidence to support differences in cognitive performance between exposed and preterm infants, at either below 12 months of age or at around 18 months of age. However, it is interesting to note that time and group main effects would have reached statistical significance at $p<.05$.

Table 8

Summary of Mental Developmental Index Means (Non-Extrapolated) for Time By Group ANOVA

		Time		
		<12 months $\overline{X}$ (SD)	18 months $\overline{X}$ (SD)	Marginal Means
Group	SHHCC	88.04 (7.63)	79.18 (10.99)	83.61
	BCCH	90.46 (5.01)	88.00 (16.90)	89.23
Marginal Means		89.25	83.59	

Table 9

Summary of ANOVA Results Using (Non-extrapolated) Mental Developmental Index Scores

Source	SS	df	MS	F	p
		Between subjects			
Group	561.540	1	561.540	4.264	.046*
Error	5135.973	39	131.692		
		Within subjects			
Time	568.695	1	568.695	6.477	.015*
Time by Group	181.573	1	181.573	2.068	.158
Error	3424.330	39	87.803		

*p<.05

When PDIex were analyzed, there were non-significant time and group main effects (refer to Table 10 and 11). However, these results need to be interpreted in light of a significant group by time interaction ($F_{1,47}$=9.510, p<.01). There is strong evidence to suggest that the rate of change in PDIex scores differed between exposed and preterm infants with the exposed infants' scores ($\overline{X}_{\text{SHHCC PDIex 1}}$ = 84.80 and $\overline{X}_{\text{SHHCC PDIex 2}}$ = 72.53) decreasing and the preterm infants' scores ($\overline{X}_{\text{BCCH PDIex 1}}$ = 72.79 and $\overline{X}_{\text{BCCH PDIex 2}}$ = 76.16) increasing over time. The interaction effect was significant despite low power (power estimated at .33).

Table 10

Summary of Psychomotor Developmental Index Means (Including Extrapolated Scores) for Time By Group ANOVA

		Time		
		<12 months $\overline{X}$ (SD)	18 months $\overline{X}$ (SD)	Marginal Means
Group	SHHCC	84.80 (7.63)	72.53 (14.65)	78.67
	BCCH	72.79 (27.11)	76.16 (32.27)	74.48
Marginal Means		78.80	74.34	

Table 11

Summary of ANOVA Results Using Psychomotor Developmental Index Scores (Including Extrapolated Scores)

Source	SS	df	MS	F	p
		Between subjects			
Group	409.030	1	409.030	415.523	.523
Error	46312.807	47	985.379		
		Within subjects			
Time	460.530	1	460.530	3.080	.086
Time by Group	1421.836	1	1421.836	9.510	.003**
Error	7027.144	47	149.514		

**$p < .01$

As with the analyses involving PDIex scores, there was a significant group by time interaction ($F_{1,39} = 27.676$, $p<.001$) when non-extrapolated PDI scores were used (refer to Table 12 and 13). Again, the exposed infants' motor scores decreased ($\overline{X}_{\text{SHHCC PDI 1}} = 88.41$ and $\overline{X}_{\text{SHHCC PDI 2}} = 76.22$) and the preterm infants' motor scores increased ($\overline{X}_{\text{BCCH PDI 1}} = 86.14$ and $\overline{X}_{\text{BCCH PDI 2}} = 98.00$) over time. However when PDI scores were analyzed, there was also a significant group main effect providing strong evidence that preterm infants, in general, outperformed exposed infants on the Motor Scale ($\overline{X}_{\text{SHHCC PDI}} = 82.31$ and $\overline{X}_{\text{BCCH PDI}} = 92.07$; $F_{1,39} = 13.154$, $p<.01$).

Table 12

Summary of Psychomotor Developmental Index Means (Non-Extrapolated) for Time By Group ANOVA

		Time		
		<12 months $\overline{X}$ (SD)	18 months $\overline{X}$ (SD)	Marginal Means
Group	SHHCC	88.41 (11.78)	76.22 (9.00)	82.31
	BCCH	86.14 (13.18)	98.00 (8.72)	92.07
Marginal Means		87.28	87.11	

Table 13

Summary of ANOVA Results Using (Non-extrapolated) Psychomotor Developmental Index Scores

Source	SS	df	MS	F	p
		Between subjects			
Group	1755.239	1	1755.239	13.154	.001**
Error	5204.005	39	133.436		
		Within subjects			
Time	.496	1	.496	.005	.943
Time by Group	2664.594	1	2664.594	27.676	<.001***
Error	3754.894	39	96.279		

**p <.01. ***p < .001

Clinically significant change was evaluated using the criterion set at a difference of 7.5 index points (or one-half SD) in mean scores from below 12 months to around 18 months of age. The exposed infants had a clinically significant decrease in cognitive and motor scores regardless of whether extrapolated scores were included. However, for the preterm infants the results were inconsistent for extrapolated and non-extrapolated scores. Preterm infants experienced a clinically significant decrease in MDIex, and increase in PDI scores, over time.

b) Change in Performance Classifications

Use of extrapolated versus non-extrapolated index scores did not effect change in performance classifications because all index scores equal to, or less than, 69 were classified as significantly delayed performance. Performance classifications were assigned a numeric value in ascending order (i.e. significantly delayed =1, mildly delayed = 2, within normal limits = 3 and accelerated performance = 4). Amount of change in classification rank was determined by subtracting the assigned value for the performance classification at time 1 from that of time 2. Forty-three percent of the SHHCC group, and 53% of the BCCH group dropped one or two rank(s) in the mental performance classification, suggesting an increased number of children would have been identified as having a possible delay in cognitive development by 18 months of age (refer to Table 14). Change in motor performance classifications differed between the two groups with 53.3% of the SHHCC group dropping in rank, while 36.8% of the BCCH group increased in rank.

Table 14

Type and Amount of Change in Performance Classifications

Change in	SHHCC				BCCH			
Performance	Mental		Motor		Mental		Motor	
Classification	n	%	n	%	n	%	n	%
-2	3	10.0	5	16.7	4	23.5	0	0
-1	10	33.3	11	36.7	5	29.4	1	5.3
0	15	50.0	11	36.7	5	29.4	11	57.9
+1	2	6.7	3	10.0	3	17.6	5	26.3
+2	0	0	0	0	0	0	2	10.5
Total	30	100	30	100	17	100	19	100

When test scores were converted to 95% CIs, the resultant test score range frequently overlapped into two performance classifications. Therefore, performance classification ranges were developed and assigned numeric ranks as follows: significantly delayed = 1, significant to mildly delayed = 2, mildly delayed = 3, mildly delayed to normal = 4, within normal limits = 5 and, normal to accelerated = 6. None of the test score ranges fell strictly in the mildly delayed performance classification range, however, the mildly delayed classification range was maintained in order to keep the classification ranks fairly equidistant (refer to Table 15).

Table 15

Type and Amount of Change in Performance Classification Using 95% Confidence Intervals

Change in	SHHCC				BCCH			
Performance	Mental		Motor		Mental		Motor	
Classification	n	%	n	%	n	%	n	%
-4	0	0	1	3.3	0	0	0	0
-3	1	3.3	1	3.3	4	23.5	0	0
-2	6	20.0	10	33.3	2	11.8	0	0
-1	6	20.0	6	20.0	4	23.5	2	10.5
0	15	50.0	8	26.7	3	17.6	7	36.8
1	1	3.3	1	3.3	3	17.6	8	42.1
2	1	3.3	3	10.0	1	5.9	1	5.3
3	0	0	0	0	0	0	1	5.3
Total	30	100	30	100	17	100	19	100

Results using 95 % CIs were similar to the results without the use of CIs. Forty-three percent of the SHHCC group and 59% of the BCCH group dropped in mental performance classification. Sixty percent of the SHHCC group decreased, while 53% of the BCCH group increased, in motor performance classification. Using 95% CI in the interpretation of change in performance classifications may not be very useful. Although reliability of the test scores is increased, the clinical meaningfulness of the changes in

classification ranges is questionable. Interpretation of change in test scores using 95% CIs at an individual, versus a group level, could be more meaningful.

Question 3: Exploration of Relationships Between Risk Variables and Test Scores

It was hypothesized that (1) biologic risk variables would be related to PDI scores and environmental risk variables to MDI scores and, (2) biologic risk variables would be more related to scores under 12 months of age than scores around 18 months of age while environmental risk variables would have the opposite effect. Null hypotheses were that there were no relationships between risk variables and high-risk infants' BSID-II scores, H_0: $\rho = 0$ (where ρ represents the linear relationships)

Relationships were explored by testing Pearson correlations between groups, birthweight, gestational age, gender, maternal age, and maternal race, with MDI 1, MDI 2, PDI 1, PDI 2, MDIex 1, MDIex 2, PDIex 1, PDIex 2, change in MDI (ΔMDI), change in PDI (ΔPDI), change in MDI with extrapolated scores (ΔMDIex) and, change in PDI with extrapolated scores (ΔPDIex). Group was coded such that BCCH group =1 and SHHCC group =2. Gender was coded such that male =1 and female = 2. Lastly, maternal ethnicity was coded such that Caucasian =1 and Non-Caucasian =2. The large number of tests run is indicative of the exploratory nature of these investigations. As well, for exploratory purposes alpha level was set at .05.

Only a few significant correlations were found (refer to Tables 16 and 17). Group status correlated significantly with MDI 2 ($r = -0.313$, $p<.05$), and PDI 2 ($r = -0.753$, $p<.001$), providing evidence that there is a relationship between group status and cognitive and motor performance at around 18 months of age (i.e. preterm infants

outperformed exposed infants). As with the ANOVA results, group status was also significantly correlated with ΔPDI ($r = -0.645$, $p<.001$) and, ΔPDIex ($r = -0.410$, $p<.01$) suggesting prenatal drug exposure was related to greater decline in motor scores over time compared to prematurity.

Maternal age was related directly to MDI 2 ($r =0.325$, $p<.05$) and ΔMDI ($r =0.333$, $p<.05$) suggesting a linear relationship between increasing maternal age and increasing cognitive scores over time. Also, there is evidence to suggest that there is a relationship between older mothers and higher cognitive scores at around 18 months of age. However, it should be noted that these relationships are small, and maternal age accounts for only about 10% of the variance in MDI 2 and ΔMDI. Maternal age was related inversely to PDI 1 ($r = -0.317$, $p<.05$), but directly to ΔPDI ($r = 0.437$, $p<.01$) and ΔPDIex ($r =0.297$, $p<05$). There is evidence to suggest a relationship between older mothers and lower motor scores in the first year of life, but greater increases in motor scores over time. Lastly, there was an isolated case of a significant correlation of PDIex 1 with birthweight ($r = 0.385$, $p<.01$) and gestational age ($r =0.318$, $p<.05$), suggesting a linear relationship between higher birthweight, larger gestational age, and higher motor scores before the first year of life. Caution must be used when interpreting these results due to the large number of tests performed.

Table 16

Correlations Between Risk Variables and Mental Scale Scores

Risk Variable		MDI 1	MDI 2	ΔMDI	MDIex 1	MDIex 2	ΔMDIex
Group	r	-.051	-.313*	-.234	.056	.096	.079
	N	51	42	38	53	47	47
Birthweight	r	.013	.007	.043	.190	.209	.117
	N	51	42	38	53	47	47
Gestational	r	.046	-.114	-.042	.166	.219	-.048
Age	N	50	42	38	52	46	48
Gender	r	.253	.020	-.113	.264	.088	-.079
	N	51	42	38	53	47	47
Maternal Age	r	-.025	.325*	.333*	-.153	-.003	.116
	N	48	39	36	50	44	44
Maternal Race	r	-.079	-.169	-.181	-.213	-.132	-.049
	N	46	39	35	48	44	44

Note. MDI 1 = Mental Developmental Index scores $\geq$ 50 at time 1; MDI 2 = Mental Developmental Index scores $\geq$ 50 at time 2; ΔMDI = change in non-extrapolated Mental Developmental Index scores; MDIex 1 = Mental Developmental Index scores including scores< 50 at time 1; MDIex 2 = Mental Developmental Index scores including scores< 50 at time 2; ΔMDIex = change in Mental Developmental Index scores including scores < 50.

*p <.05

Table 17

Correlations Between Risk Variables and Motor Scale Scores

Risk Variable		PDI 1	PDI 2	ΔPDI	PDIex 1	PDIex 2	ΔPDIex
Group	r	.101	-.753***	-.645***	.252	-.068	-.410**
	N	49	42	39	53	49	49
Birthweight	r	.260	.015	-.218	.385**	.263	-.115
	N	49	42	39	53	49	49
Gestational	r	.104	-.191	-.185	.318*	.265	-.048
Age	N	48	42	39	52	48	48
Gender	r	-.040	-.055	-.051	.129	.048	-.075
	N	49	42	39	53	49	49
Maternal Age	r	-.317*	.216	.437**	-.193	.065	.297*
	N	46	39	36	50	46	46
Maternal Race	r	-.082	-.059	.075	-.272	-.179	.073
	N	44	39	36	48	46	46

Note. PDI 1 = Psychomotor Developmental Index scores $\geq$ 50 at time 1; PDI 2 = Psychomotor Developmental Index scores $\geq$ 50 at time 2; ΔPDI = change in non-extrapolated Psychomotor Developmental Index scores; PDIex 1 = Psychomotor Developmental Index scores including scores< 50 at time 1; PDIex 2 = Psychomotor Developmental Index scores including scores< 50 at time 2; ΔPDIex = change in Psychomotor Developmental Index scores including scores < 50.

*p <.05. **p < .01. ***p < .001

Only the variables that were significantly related to the outcome variables were used as predictor variables in the MLR analyses. MLR analyses were performed from a minimum of two predictors to a maximum of five. Only the dependent variables MDI 2, ΔPDI, and ΔPDIex had significant correlations with more than one variable (i.e. maternal age and group status). Group ($R=0.338$; $t_{1,37}= -2.183$, $p<.05$) was the only variable that entered into the equation to predict MDI 2 (refer to Table 18). Only nine percent of the variance in MDI 2 scores could be predicted by group status (adjusted $R^2= 0.09$). The resultant equation was Y = 97.769 – 9.769 * Group (where Y is the predicted MDI score at 18 months, preterm infant group =1 and exposed infant group =2). The equation can predict MDI 2 scores within 26.36 index points 99% of the time. The error associated with the prediction is almost 2 SDs large. Therefore, predictions using the equation were deemed not useful.

Table 18

Summary of Regression Analysis for Variables Predicting Non-Extrapolated Mental Developmental Index Scores at the Second Test Time

Model	B	SE B	β	t	p
(Constant)	97.769	7.753		12.611	<.001***
Group	-9.769	4.476	-.388	-2.183	.035*

Note. $R=0.338$; $R^2=0.114$; adjusted $R^2=0.090$; Standard Error of the Estimate = 13.18. Maternal age was an excluded variable (i.e. did not enter into the equation).

*p<.05. ***p<.001

Both group and maternal age entered into the equation to predict ΔPDI (refer to Table 19). First, group status entered the equation accounting for about 43% (R^2=.0425) of the variance. Second, maternal age entered the equation accounting for approximately 5% (Δ R^2=0.053) more of the variance in ΔPDI. Together, after adjustments, 45% of the variance in ΔPDI could be predicted by the variance in group status and maternal age (adjusted R^2=0.446; $F_{2,33}$= 15.108, p<.001). The resultant equation for this sample was Y = 9.481 – 19.721 * Group + 0.711 * Maternal Age (where Y is the predicted value of ΔPDI, preterm infants = 1 and exposed infants = 2). Taking into account the standard error of estimate, change in PDI can be predicted by group status and maternal age within 25.48 index points approximately 99% of the time. However, an index point spread so large (i.e. greater than one and one-half SDs) is not sensitive enough to be considered clinically meaningful. Furthermore, a great deal of caution must be used in generalizing this equation as cross-validation with other samples has not been done.

Table 19

Summary of Regression Analysis for Variables Predicting Change in Non-Extrapolated Psychomotor Developmental Index Scores

Model	B	SE B	β	t	p
Step 1					
(Constant)	34.442	7.579		4.544	<.001***
Group	-22.584	4.503	-.652	-5.016	<.001***
Step 2					
(Constant)	9.481	15.513		.611	.545
Group	-19.721	4.630	-.569	-4.260	<.001***
Maternal Age	.711	.390	-.244	1.826	.077

Note. At Step 1 R=0.652; R^2=0.425; adjusted R^2=0.408; Standard Error of the Estimate = 13.17. At Step 2 R=0.691; R^2=0.478; adjusted R^2=0.446; Standard Error of the Estimate = 12.74; ΔR^2=0.053.

***p<.001

Group (R=0.390; $t_{1,50}$= -2.806, p<.01) was the only variable that entered into the equation to predict change in ΔPDIex (refer to Table 20). Only 13% of the variance in ΔPDIex could be predicted by the variance in group status (adjusted R^2=0.133). The resultant equation obtained was Y = 17.515 – 14.146 * Group (whereY is the predicted value of ΔPDIex, preterm infant group = 1, and exposed infant group = 2). The equation

could predict ΔPDIex within 33.68 points 99% of the time. Again, the error associated with the predictions is very high.

Table 20

Summary of Regression Analysis for Variables Predicting Change in Extrapolated Psychomotor Developmental Index Scores

Model	B	SE B	β	t	p
(Constant)	17.515	8.377		2.091	.042*
Group	-14.146	5.041	-.390	-2.806	.007**

Note. R=0.390; R^2=0.152; adjusted R^2=0.133; Standard Error of the Estimate =16.84. Maternal age was an excluded variable (i.e. did not enter into the equation).

$p<.05$. **$p<.01$

Overall there were only a few, mostly isolated cases of significant correlations between risk variables and outcome variables. As well, there were only a few cases where more than one risk variable related significantly to an outcome variable. When linear regression analyses were performed, ΔPDI was the only outcome measure where more than one risk variable entered into the regression equation. Errors associated with the regression equations were high, and therefore, not clinically useful. There is little evidence (using this sample) that biological and environmental risk variables can be used to make accurate predictions in motor and cognitive developmental outcomes as measured by the BSID-II.

CHAPTER 5

Discussion

This chapter will summarize the results of each of the research questions. These results will then be compared to the results reported in the literature. Limitations of this study, including future directions for research and practice, will be addressed.

Identification of Developmental Delays

Performance by infants prenatally exposed to alcohol and/or other drugs (exposed infants) and by infants born prematurely and/or with significant perinatal medical concerns (preterm infants) on the Bayley-II Mental and Motor Scale was significantly lower at under one year of age, and around 18 months of age, compared to the normative mean. Statistical significance was obtained when using either extrapolated or non-extrapolated scores in the analyses. The means obtained were also rated to be clinically significantly different from the norms using the set criterion of a difference of 7.5 index points, or one-half a SD. The percentage of infants who were classified as having a significant, or mild, motor or cognitive delay using the BSID-II performance classifications was larger than what would be expected by a group of typically-developing children (i.e. percentage of children who would score below one SD). Therefore, this study adds to the clinical validity of the BSID-II for the identification of high-risk infants.

Extrapolated scores were not used in any of the published studies using the BSID-II. Therefore, non-extrapolated scores were used for comparing the results from this study to those previously published. When the published study had only one testing time, the

results were compared to this study's scores obtained at the most similar testing age. Statistical analyses comparing results between this study and other published studies were not performed. Instead, differences in performance were evaluated using the criterion established for a clinically significant difference of 7.5 index points, or one-half a SD. Corresponding to the literature, the results of this study were divided into the two groups (i.e. infants prenatally exposed to alcohol and/or drugs and, infants born prematurely and/or with significant perinatal medical concerns).

Infants Prenatally Exposed to Alcohol and/or Drugs

The study presented in the BSID-II manual on children who had been prenatally exposed to drugs reported a higher MDI mean when compared to this study's exposed group mean at 18 months of age. The BSID-II manual's study obtained higher PDI means than those obtained in this study at both testing times (refer to Table 19). The differences in the results may have been due to this study sample's younger age at testing. However, it appears more likely to be due to the nature of this study's sample. In the present study, all but one child was admitted to hospital for treatment of Neonatal Abstinence Sydrome. Therefore, the children were not only exposed to drugs prenatally, but their central nervous system was affected by the exposure. In other words, this sample of children may have been at greater biologic risk at birth, and hence obtained lower Bayley-II scores.

Some similarities in results were found between this study and the study by Alessendri and colleagues (1998). Although this study's exposed infants appeared to perform more poorly on the Mental Scale at the first testing compared to the high cocaine exposure group, performance was similar at the second testing time (refer to Table 19).

This sample's performance on the Motor Scale appeared to be similar to both the high cocaine and the low exposure groups at time one. The few differences in the results between this and the Alessendri et al. study may have been due to the differences in sampling. In this study, the infants were exposed to a mixture of drugs, not necessarily predominantly to cocaine.

Table 21

Summary of Study BSID-II Means Obtained with Children Prenatally Exposed to Drugs

Study	Sample	Age(s)	MDI $\overline{X}$ (SD)	PDI $\overline{X}$ (SD)
Present Study	N=19 exposed to alcohol and/or other drugs	>12 mo. 18 mo.	86.97 (8.65) 79.00 (10.83)	85.12 (13.62) 75.36 (9.95)
BSID-II manual (1993)	N=137 with polydrug exposure	Median age = 24 mo. (1 to 40 mo.)	90.8 (16.2)	96.3 (20.0)
Alessandri et al. (1998)	N=15 with high exposure to cocaine N=19 with low exposure to cocaine	8 mo. 18 mo. 8 mo. 18 mo.	94.59 (8.91) 79.05 (10.21) 91.30 (7.05) 86.59 (9.76)	87.33 (9.93) N/A 89.99 (8.70) N/A
Heffelfinger et al. (1997)	N=14 with cocaine exposure	Mean age 20 mo. (8 to 40 mo.)	86.64 (12.00)	N/A

Lastly, this study's exposed infants appeared to perform more poorly on the Mental Scale, especially at around 18 months, compared to the sample used in Heffelfinger and colleagues' study in 1997 (refer to Table 19). This difference may have been due to the wider age range used in the Heffelfinger et al. study and due to the differences in type of prenatal drug exposure.

None of the published studies that used the BSID-II with infants prenatally exposed to drugs converted test scores into performance classifications. Therefore, comparison of incidence of performance classifications obtained in the present study could not be made.

In summary, when there were differences found in the results between the present study and the studies published in the literature, this study's exposed infants appeared to perform more poorly. The poorer performance on the BSID-II by this sample, especially before 12 months of age, could possibly be due to the higher perinatal biological risk associated with Neonatal Abstinence Syndrome. However, the possible lower environmental risk associated with foster care environments for this sample did not result in higher cognitive performance at 18 months of age.

Infants Born Prematurely

The results at under 12 months of age for this study's preterm infants were similar to results of the study presented in the BSID-II manual (1993) with children born prematurely for both the Mental and Motor Scales (refer to Table 20). These results were similar despite the differences in the definitions of prematurity (i.e. $\leq$ 37 weeks GA in this study and < 36 weeks GA in the Bayley study), gender ratios (i.e. 73.7% males in this

study and 37% males in the Bayley study), and severity of perinatal medical concerns (i.e. possibly more severe in this study than in the Bayley study).

This study's preterm infants appeared to perform more poorly on both the Mental and Motor Scales at under one year of age compared to the 1998 study by Case-Smith and colleagues (refer to Table 20). The lower scores obtained by this study's preterm infants may have been due to the younger age at testing, when the effects of perinatal medical concerns are believed to be more influential.

Although the mean age of testing in the study by D.J. Goldstein and colleagues (1995) was more similar to this study's first testing age, the range of ages was more reflective of the second testing age (refer to Table 20). Therefore, results from both testing times were compared. This study's preterm infants appeared to perform similarly on the Mental Scale at both testing times compared to the D.J. Goldstein et al. sample. Also, motor performance was similar when compared to this study's performance at below 12 months of age. At around 18 months of age, this study's preterm infants appeared to outperform the D.J. Goldstein et al. sample on the Motor Scale. The better motor performance obtained may be explained by the "catch-up" phenomenon often experienced by preterm infants with milder perinatal medical concerns later in life.

Lastly, the means obtained from this study were compared to those obtained in the 1998 study by Macais and colleagues (refer to Table 20). The Macais et al. sample appeared to obtain slightly higher MDI scores compared to this study's preterm infants' scores at both testing times. Although the age range tested in the Macais et al. study was greater and their sample may have been more heterogeneous (e.g. inclusion of children

with prenatal drug exposure) compared to the present study sample, the results obtained were similar. The Motor Scale was not administered in the Macais et al. study.

Table 22

Summary of Study BSID-II Means Obtained with Children Born Prematurely

Study	Sample	Age(s)	MDI $\overline{X}$ (SD)	PDI $\overline{X}$ (SD)
Present Study	N=19 preterm and/or medical concerns	>12 mo. 18 mo.	87.89 (8.81) 88.00 (16.90)	82.00 (16.69) 98.00 (8.72)
BSID-II manual (1993)	N= 57 preterm, some with medical concerns	Median age = 11 mo. (2 to 27 mo.)	88.6 (15.7)	83.5 (21.6)
Case-Smith et al. (1998)	N= 45 preterm with medical concerns	12 mo.	97.74 (15.5)	89.90 (20.8)
D.J Goldstein et al. (1995)	N=37 preterm with medical concerns	Mean age = 12.8 mo. (11 to 20 mo.)	92.77 (15.80)	83.00 (16.55)
Macais et al. (1998)	N=78 preterm and/or medical concerns (n=6 with prenatal cocaine exposure)	Mean age = 12.9 mo. (6 to 24 mo.)	91.6 (17)	N/A

In a study by Doig and colleagues (1999) mean scores were not reported; however the percentages of their sample who scored below 1 and 2 SDs from the normative mean (i.e. classified as having a mild or significant delay using the BSID-II performance classifications) were reported. Two other studies also reported percentages of children who would have been classified with a mild or significant developmental delay (D.J. Goldstein et al., 1995; Macais et al.,1998). None of the studies converted test scores to 95% CIs prior to classifying performance. Therefore, the results from this study's preterm group without 95% CIs conversions will be used for all comparisons.

Doig and colleagues (1999) administered the BSID-II Mental Scale to 36 infants at a mean age of 25.5 months (ranging from 15 to 40 months of age). Nineteen percent of the children were classified as having a mild delay in mental performance, while 28 % were classified as having a significant delay. When comparing these results to those obtained in the present study at around 18 months of age, more of this study's sample was classified as having a significant delay (41.2%). The percentage of infants identified with a mild delay in cognitive performance was similar at 17.6%.

It is unclear why this study's sample had a higher percentage of infants classified with a significantly delay compared to the study by Doig and colleagues (1999). From the sample description, it does not appear that this study's preterm infants necessarily had more severe perinatal medical concerns. Furthermore, the Doig et al. sample had a lower average GA of 25.5 weeks compared to this sample ($\overline{X}_{GA}$ = 32.21 weeks). Perhaps the difference in results was due the difference in the definition of prematurity between the studies. Corrected ages were used for infants born at a GA $\leq$ 37 weeks in this study and

GA < 36 weeks in the Doig et al. study. Using corrected versus chronological ages for testing on the BSID-II can lead to lower scores (Ross & Lawson, 1997). However, it is doubtful whether the slight differences in prematurity definitions would lead to such large differences in performance classification, especially for the significantly delayed classifications. The differences in the percentage of significantly delayed classification may have been due to the inclusion of all index scores, even those below 50, in the classification of performance.

D.J. Goldstein and colleagues (1995) reported that 27% of their sample obtained MDI scores which were classified as being in the mild or significantly delayed range, whereas 49% of the PDI scores were in that range. D.J. Goldstein and colleagues did not report separate percentages of children classified with a mild or significant delay. The present study obtained higher percentages of infants that were classified as having either a mild or significant delay at both testing occasions (MDI 1 = 31.6%, MDI 2 = 58.8%, PDI 1 = 57.9%, and PDI 2 =31.6 %). Perhaps fewer subjects in the Goldstein et al. sample were classified as having a cognitive or motor delay because subjects with major motor abnormalities were excluded.

Macais and colleagues (1998) reported the performance classifications for 78 children tested on the BSID-II Mental Scale. Eleven and one-half percent of their sample was classified with a mild cognitive delay and 14.1% with a significant delay. The results from the Macais et al. study are comparable to the results of the present study at the first testing occasion (below 12 months of age) with 10.5% of the infants in this study classified with a mild delay and 21% with a significant delay. Alternatively, if the Macais study results are compared to this study's second testing time (around 18 months of age),

more of this study's sample was classified with a significant delay (41.2 %), whereas the incidence of a mild delay was similar (17.6%). The differences in these results may be due to the differences in age at testing. Although children born prematurely with less severe perinatal medical concerns appear to "catch up" in their physical development, this may not necessarily apply to their cognitive development. Because cognitive expectations and different types of thinking (i.e. problem solving) increase with age, perhaps more children are identified at 18 months. However, this trend did not appear in the literature review of studies using the original Bayley Scales. It is more probable that the differences in performance classification arose due to differences in sampling.

In summary, there were a few differences in the performance of preterm infants between the present study and other published studies using the BSID-II. However, it is not clear if these differences are significant. Direct comparison of the different study results is difficult due to the differences in the sampling (e.g. inclusion of infants with physical disabilities, severity of perinatal medical concerns, age at testing, etc.). Lack of consistency in definitions of prematurity based on GA further contributed to the problem. Not only do these varied definitions change the type of sample recruited, but also the problem is accentuated by the BSID-II itself. Using chronological versus corrected age item sets on the BSID-II can lead to lower scores (Ross & Lawson, 1997). Although all studies used corrected ages up to 2 years of age when administering and scoring the Bayley II, using children 24 to 40 months of age in the sample could have an effect on the overall outcome for the sample. Once more studies are published using the BSID-II with infants born prematurely and/or with significant perinatal medical concerns, a meta-analysis of the results could be useful.

Change in Performance Over Time

Use of extrapolated index scores in the analyses yielded some different results than use of only non-extrapolated index scores. Statistical analyses with extrapolated scores revealed that exposed and preterm infants' cognitive performance declined from under 12 months of age to around 18 months of age. Analyses using non-extrapolated scores revealed that preterm infants had better overall motor performance than exposed infants. The only result that was consistent for both extrapolated and non-extrapolated scores was that there was a significant difference in the rate of change in motor performance. The preterm infants' motor scores increased while the exposed infants' motor scores decreased from under 12 months to around 18 months of age.

If clinical versus statistical differences are used to evaluate the results, exposed infants had a decline in both mental and motor performance over time regardless of whether extrapolated or non-extrapolated scores were used. Clinically significant change in performance of preterm infants was inconsistent. There was a clinically significant decrease in cognitive performance if extrapolated scores were used, and a clinically significant increase in motor performance if non-extrapolated scores were used.

There was only one study (Alessandri et al., 1998) and one abstract (Mattia & deRegnier,1998) collected BSID-II scores on more than one occasion. In the published studies currently available, extrapolated scores were not used. As such, the results from the present study using non-extrapolated scores were used for all comparisons. In a study by Alessandri and colleagues (1998) of infants prenatally exposed to cocaine, cognitive scores were found to decrease from 8 to 18 months of age. The non-exposed infants in their study obtained higher MDI scores at 18 months

compared to the high-cocaine exposed infants. In the present study, the change in cognitive scores over time did not reach significance with alpha level set at .01. However, if alpha level had been set at .05, then both the time and the group main effects would have reached statistical significance. MDI scores for the sample as a whole would have significantly deceased over time, and the preterm infants would have obtained significantly higher MDI scores averaged over time when compared to the exposed infants.

An abstract by Mattia and deRegnier (1998) collected MDI and PDI scores for 96 extremely premature infants (GA $\leq$ 30 weeks) at 1 year and at 2 to 3 years of age. However, neither means, nor change in test scores over time, were discussed. There were no published studies that discussed changes in performance classifications over time.

Exploration of Relationships Between Risk Variables and Performance

The present study obtained only a few, isolated instances of significant linear relationships between risk variables and BSID-II outcomes. Group status predicted cognitive performance at 18 months of age and change in motor scores over time, with preterm infants outperforming exposed infants in these instances. Maternal age was a second predictor of change in motor performance when non-extrapolated PDI scores were used in the analysis. In this case, prematurity (versus prenatal drug exposure) and having an older mother appeared to predict a greater improvement in motor scores over time. Although these few predictions were statistically significant, the explained variance in test scores was small and not considered to be clinically meaningful.

One study with infants prenatally exposed to drugs (Alessandri et al., 1998) and one abstract with infants born prematurely (Mattia & deRegnier, 1998) investigated

relationships between risk variables and BSID-II outcomes. In the study by Alessandri and colleagues, hierarchical regression analyses were conducted using 8- month MDI scores, 18- month MDI scores, change in MDI scores, and 8-month PDI scores as dependent variables. Cocaine was entered first, followed by other drug exposure, medical risk, environmental risk, and lastly amount of prenatal cocaine exposure (high, low, or none) by environmental risk. Eighteen-month MDI scores were predicted from these variables (Total $R^2 = .14$) with group status ($R^2 = .05$) and environmental risk ($R^2 = .04$) as significant independent predictors ($p<.05$). Eight-month PDI scores were also predicted from these variables ($R^2 = .17$) with group status ($R^2 = .11$) as a significant independent predictor ($p< .05$) when first entered into the equation. Therefore, there is strong evidence suggesting higher prenatal cocaine exposure and higher environmental risk led to lower cognitive scores at 18 months of age and lower motor scores at 8 months of age. The risk variables used in their analyses did not significantly predict either 8- month MDI scores nor change in MDI scores from 8- to 18- months ($p>.05$).

Mattia and deRegnier (1998) investigated relationships between biologic risk variables and BSID-II scores for infants born prematurely. They performed MLR with a measure of chronic physiologic instability, intraventricular hemorrhage (IVH) scores, gestational age (GA), and weight-change in first month as the independent variables. Higher physiologic instability scores, IVH scores, and GA were associated with lower one-year MDI scores (R-values were not reported). At two years of age only higher physiologic stability score was related to lower MDI score. Higher physiologic instability and IVH scores led to lower PDI scores at one year of age. Interestingly, these results

suggest that biologic risk factors have a more long-term effect on cognition rather than on motor development.

It was difficult to compare results from this study to those from other published studies because of the different risk variables used in the analyses. Methods and variable types used to measure environmental and biological risk differed. Furthermore, Alessandri and colleagues (1998) used outcomes from infants with varying degrees of prenatal cocaine exposure and, Mattia and deRegnier (1998) used outcomes from infants born prematurely, whereas the present study used outcomes from both infants with prenatal polydrug exposure and infants born prematurely and/or with significant prenatal medical concerns. It appears that more research is needed in this area, but foremost more universal measures of risk variables should be in place.

Limitations and Future Directions

The present study adds to the clinical validity of the BSID-II in the identification of infants at risk for developmental delays. Use of extrapolated scores for index scores that fell below 50 enabled the use of more subjects' scores in the statistical analyses. There were eight subjects for whom extrapolated scores were used (3 with prenatal drug exposure and 5 born prematurely). Among these subjects, there were a total of 18 instances for which extrapolated scores were used (2-MDI 1, 4-PDI 1, 5-MDI 2, and 7-PDI 2 scores). Of these, there were 11 instances (1-MDI 1, 2-PDI 1, 3-MDI 2, and 5-PDI 2 scores) where the formulae provided by Robinson and Mervis (1996) led to negative results and therefore, an index score of 15 was assigned. Interestingly, all eight subjects for whom extrapolated scores were used later received diagnoses (6 with cerebral palsy, 2 with global developmental delay).

The medical chart review revealed eight more subjects with later developmental difficulties (e.g. speech delays, fine motor difficulties, behavioral problems, Attention Deficit Disorder, hearing loss, low muscle tone, tremor etc.), who did not receive index scores below 50. All of these subjects scored in the mildly to significantly delayed range on either the Mental or the Motor scale by 18 months of age. Furthermore, seven of the eight subjects experienced a large drop (ranging from 8 to 27 index points) in MDI and/or PDI scores from under a year to 18 months of age. It is unknown if more of the subjects in the present study experienced later cognitive or motor difficulties, as this chart review was limited to the two sites. Both sites are tertiary care institutions, and therefore, it is likely that other children would receive direct therapy services elsewhere. Overall, it appears that infants who scored below 50 at 18 months of age would likely receive a diagnosis later in life. Other infants with less severe cognitive and physical impairments, but with difficulties in one or more areas of development (e.g. speech and language, fine motor, gross motor, behavior) appeared to experience a significant drop in scores from under one year to around 18 months of age.

Although use of extrapolated scores allowed for inclusion of all subjects in the statistical analyses, there were some problems encountered with their use. The majority of the scores (i.e. 61%) which fell below 50 obtained negative scores using the formulae provided by Robinson and Mervis (1996). Inclusion of extrapolated scores led to more skewed and variable distributions, at times violating the assumptions related to some statistical tests. Increasing the number of statistical tests conducted, so that both extrapolated and non-extrapolated scores could be analyzed, led to a more stringent alpha level and, therefore, a reduction in the power to detect 'true' difference in the population.

Increasing sample size could increase power in future studies. Use of extrapolated scores versus non-extrapolated scores resulted in different statistical and clinical interpretations when analyzing change over time, group differences, and group by time interaction effects.

Because none of the studies published thus far using the BSID-II with infants prenatally exposed to drugs and infants born prematurely have used extrapolated scores in their analyses, it may be wise to perform future statistical analyses using non-extrapolated scores only to reduce the number of tests performed. Performance classifications, which are easily derived regardless of an index score below 50, are an alternate way of including all of the subjects' scores in data analyses.

Converting index scores to performance classifications can assist clinicians in identification of infants who require therapeutic interventions. Use of 95% CIs has been recommended to increase the reliability of test scores and to assist clinicians in deciding if further testing is needed (Cunningham-Amundson, & Crowe, 1993). Use of 95% CIs resulted in test score ranges. These ranges in scores led to ranges in performance classifications. Unfortunately, use of performance classification ranges that incorporate 95% CIs made the classifications less meaningful, especially when evaluating change in performance classifications over time. Ninety-five percent CIs may be more clinically relevant when interpreting individual versus group performance.

Exploring relationships between biologic and environmental risk variables and BSID-II cognitive and motor outcomes could help shape developmental theories and guide clinical practice. In the present study, there were many limitations in the risk variables used in the analyses. Choice of risk variables, especially environmental ones,

was limited to the information available in the subjects' medical records. It is questionable whether the environmental risk variables used in this study (i.e. group status, maternal age and maternal ethnicity) were truly risk variables. It is assumed that the environmental risk is greater for infants prenatally exposed to drugs due to the risk associated with a drug using lifestyle/home environment. However in the present study, the majority of the infants in the exposed group were living with foster families. Maternal age and ethnicity are considered 'distal' risk variables because their effects on the quality of the environment are neither direct, nor easily changed. 'Proximal' environmental risk variables that measure such things as the quality of caregiving, social interactions, and stimulation, could be better indicators of environmental risk. Alessandri and colleagues (1998) used cumulative environmental risk measures to study relationships with BSID-II scores obtained by infants prenatally exposed to cocaine. An environmental risk score was cumulated by using a number of different risk factors (e.g. life stressors, social support network size, maternal educational level, minority status, and number of children in the household). Use of both distal and proximal risk variables in a cumulative fashion could be a more accurate overall indicator of environmental risk. The same may be true when attempting to measure biologic risk. In the study by Mattia and deRegnier (1998), a cumulative measure of physiological instability was used, as well as IVH scores, GA, and change in first month weight. Although cumulative biologic and environmental risk scores could not be used in this retrospective chart review, this should be a consideration for future studies. How, and what variables to include, to measure biologic and environmental risk require further debate and research. Until more universally accepted

risk variables are used to study relationships and to predict cognitive and motor outcomes, it will continue to be difficult to compare results from different studies.

In summary, infants at high-risk for developmental delays scored lower on the BSID-II compared to typically-developing infants. Use of performance classifications, along with clinical observations, could assist in the identification of developmental delays. Also, performance classifications allow for the inclusion of all subjects in the analyses, even these with MDI and PDI scores below 50. More studies are needed to measure change in cognitive and motor performance over time. Infants prenatally exposed to alcohol an/or other drugs experienced a decrease in motor scores while infants born prematurely and/or with significant perinatal medical concerns experienced an increase in motor scores from under one year to 18 months. The decrease in motor scores for the infants exposed to drugs was not expected and warrants further investigation. Alternatively, a decrease in cognitive scores was expected for exposed infants over time. If alpha level had been set at .05 as opposed to .01, the decrease in exposed infants' MDI scores would have reached statistical significance. Prospective longitudinal studies are needed in order to verify these trends. A larger sample, as opposed to a more stringent alpha level, could be a better way to increase the power of analyses in future studies. Studies with a greater number of testing times, especially at older ages, would also be useful. Lastly, more research is needed in measurement of biologic and environmental risk variables. Although it appears that cumulative risk scores may be better indicators of risk than individual variables, which variables to use and how to measure them needs further investigation and consistency.

References

Alessandri, S.M., Bendersky, M., & Lewis, M. (1998). Cognitive functioning in 8- to 18-month-old drug-exposed infants. Developmental Psychology,34(3), 565-573.

Anderson, A.E, Wildin, S.R., Woodside, M., Swank, P., Smith, K.E., Denson, S.E., Miller, C.L., Butler, I.J., & Landry, S.H. (1996). Severity of medical and neurologic complications as a determinant of neurodevelopmental outcome at 6 and 12 months in very low birth weight infants. Journal of Child Neurology, 11, 215-219.

Aylward, G.P., Pfeiffer, S.I., Wright, A., & Verhulst, S.J. (1989). Outcome studies of low birth weight infants published in the last decade: A metaanalysis. Journal of Pediatrics, 115(4), 515-520.

Aylward, G.P., Gustafson, N., Verhulst, S.J. & Colliver, J.A. (1987). Consistency in the diagnosis of cognitive, motor, and neurological function over the first three years. Journal of Pediatric Psychology, 12(1), 77-98.

Bayley, N. (1969). Bayley scales of infant development. New York, NY: Psychological Corporation.

Bayley, N. (1993). Bayley scales of infant development second edition: Manual. San Antonio: Psychological Corporation.

Bendersky, M., & Lewis, M. (1994). Environmental risk, biological risk, and developmental outcome. Developmental Psychology, 30(4), 484-494.

Billman, D.O., Nemeth, P.B., Hiemler, R., & Sasidharan, P. (1996). Prenatal cocaine/ploydrug exposure: Effect of race on outcome. Journal of Perinatology, 16(5), 366-369.

Bornstein, M.H., & Sigman, M.D. (1986). Continuity in mental development from infancy. Child Development, 57, 251-274.

Brady, D.K., Crowe, T.K., & Deitz, J.C. (1992). Developmental outcome in appropriate and small for gestational age premature infants. Pediatric Physical Therapy, 5, 18-23.

Brazy, J.E., Eckerman, C.O., Oehler, J.M., Goldstein, R.F., & O'Rand, A.M. (1991). Nursery neurobiologic risk score: Important factors in predicting outcome in very low birth weight infants. Journal of Pediatrics, 118(5), 783-792.

Brazy, J.E., Goldstein, R.F., Oehler, J.M., Gustafson, K.E., & Thompson, Jr., R.J. (1993). Nursery neurobiologic risk score: Levels of risk and relationship with nonmedical factors. Journal of Developmental and Behavioral Pediatrics, 14(6), 375-380.

Brooks-Gunn, J., Klebanov, P.K., Liaw, F.R, & Spiker, D. (1993). Enhancing the development of low-birthweight, premature infants: Changes in cognition and behavior over the first three years. Child Development, 64, 736-753.

Caldwell, B.M., & Bradley, R.H. (1979). Home Observation for Measurement of the Environment. Little Rock, Arkansas: University of Arkansas Press.

Campbell, S.K., Siegel, E., Parr, C.A. & Ramey, C.T. (1986). Evidence for the need to renorm the Bayley Scales of Infant Development based on the performance of a population-based sample of 12-month-old infants. Topics in Early Childhood and Special Education, 6(2), 83-96.

Case-Smith, J., Butcher, L., & Reed, D. (1998). Parents' report of sensory responsiveness and temperament in preterm infants. American Journal of Occupational Therapy, 52(7), 547-555.

Chasnoff, I.J. (1991). Cocaine and pregnancy: Clinical and methodological issues. Clinics in Perinatology, 18(1), 113-123.

Chasnoff, I.J., Burns, K.A., Burns, W.J., & Schnoll, S.H. (1986). Prenatal drug exposure: Effects on neonatal and infant growth and development. Neurobehavioral Toxicology and Teratology,8(4), 357-362.

Chasnoff, I.J., Griffith, D.R., Freier, C., & Murray, J. (1992). Cocaine/polydrug use in pregnancy: Two-year follow-up. Pediatrics, 89(2), 284-289.

Cherny, S.S., Fulker, D.W., Emde, R.N., Robinson, J., Corley, R.P., Resnick, J.S., Plomin, R., & DeFries, J.C. (1994). A developmental-genetic analysis of the continuity and change in Bayley Mental Development Index from 14 to 24 months: The Mac Arthur longitudinal twin study. Psychological Science, 5(6), 354-360.

Cole, K.N., & Harris, S.R. (1992). Instability of the intelligence quotient - motor quotient relationship. Developmental Medicine and Child Neurology,34, 633-641.

Cooper, P.A., & Sandler, D.L. (1997). Outcome of very low birth weight infants at 12 to 18 months of age in Soweto, South Africa. Pediatrics,99(4), 537-544.

Corcoran, J.D., Patterson, C.C., & Halliday, H.L. (1993). Reduction in risk of bronchopulmonary dysplasia from 1980-1990: Results of a multivariate logistic regression analysis. European Journal of Pediatrics,152, 677-681.

Coryell, J., Provost, B., Wilhelm, I.J., & Campbell, S.K. (1989). Stability of the Bayley motor scale scores in the first year of life. Physical Therapy, 69(10), 834-841.

Cosden, M., Peerson, S., & Elliot, K. (1997). Effects of prenatal drug exposure on birth outcomes and early childhood development. Journal of Drug Issues,27(3), 525-539.

Costarides, A.H. & Shulman, B.B. (1998). Norm-referenced language measures: Implications for assessments of infants and toddlers. Topics in Language Disorders, 18(3), 26-33.

Cunningham-Amundson, S.J., & Crowe, T.K. (1993). Clinical application of the standard error of measurement for occupational and physical therapists. Physical and Occupational Therapy in Pediatrics, 12(4), 57-71.

Darrah, J., Redfern, L., Maguire, T.O., Beaulne, P., & Watt, J. (1998). Intra-individual stability of rate of gross motor development in full-term infants. Early Human Development, 52, 169-179.

Day, N.L. & Richardson, G.A. (1991). Prenatal marijuana use: Epidemiology, methodologic issues, and infant outcome. Clinics in Perinatology, 18(1), 77-91.

Day, N.L., Richardson, G.A. & McGauhey, P.J. (1994). The effects of marijuana, cocaine, heroine, and methadone. In H.L. Needleman & D. Bellinger (Eds.), Prenatal Exposure to Toxicants: Developmental Consequences, (pp. 184-209). Baltimore, MA: John Hopkins University Press.

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 the follow-up of the high-risk infant. Journal of Pediatrics, 135(3), 358-362.

Domholdt, E. (1993). Physical therapy research: Principles and applications. Philadelphia, PA: W.B. Saunders Company.

Dunn, W.W. (1989). Validity. Physical and Occupational Therapy in Pediatrics, 9(1), 149-168.

Dunst, C.J. & Rheingrover, R.M. (1981). Discontinuity and instability in early development: Implications for assessment. Topics in Early Childhood Special Education, 1(2), 49-60.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post hoc, and compromise power analyses for MS-DOS [Computer Program]. Bonn, FRG: Bonn University, Department of Psychology.

Favell, J.H., Miller, P.H., & Miller, S.A. (1993). Cognitive development (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall.

Feingold, C. (1994). Correlates of cognitive development in low-birth-weight infants from low-income families. Journal of Pediatric Nursing: Nursing Care of Children and Families, 9(2), 91-97.

Flanagan, D.P. & Alfonso, V.C. (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. Journal of Psychoeducational Assessment, 13, 66-90.

Fried, P.A., & Watkinson, B. (1988). 12- and 24-month neurobehavioural follow-up of children prenatally exposed to marihuana, cigarettes, and alcohol. Neurotoxicology and Teratology, 10(4), 305-313.

Gauthier, S.M., Bauer, C.R., Messinger, D.S., & Closius, J.M. (1999). The Bayley Scales of Infant Development II: Where to start? Journal of Developmental and Behavioral Pediatrics, 20(2), 75-79.

Gennaro, S., & Stringer, M. (1991). Stress and health in low birthweight infants: A longitudinal study. Nursing Research, 40(5), 308-311.

Glass, G.V., & Hopkins, K.D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

Glenberg, A.M. (1988). Learning from data: An introduction to statistical reasoning. San Diego: Harcourt Brace Javanovich.

Golden, N.L., Sokol, R.J., Kuhnert, B.R., & Bottoms, S. (1982). Maternal alcohol use and infant development. Pediatrics, 70(6), 931-934.

Goldstein, D.J., Fogle, E.E., Wieber, J.L., & O'Shea, T.M. (1995). Comparison of the Bayley Scales of Infant Development-second edition and the Bayley Scales of Infant Development with premature infants. Journal of Psychoeducational Assessment, 13, 391-396.

Goldstein, R.F., Thompson, Jr., R.J., Oehler, J.M., Brazy, J.E. (1995). Influence of acidosis, hyperemia, and hypotension on neurodevelopmental outcome in very low birth weight infants. Pediatrics, 95(2), 238-243.

Gross, S.J., Slagle, T.A., D'Eugenio, D.B., & Mettelman, B.B. (1992). Impact of a matched control group on interpretation of developmental performance in preterm infants. Pediatrics, 90(5), 681-687.

Gusella, J.L., & Fried, P.A. (1984). Effects of maternal social drinking and smoking on offspring at 13 months. Neurobehavioral Toxicology and Teratology, 6(1), 13-17.

Hans, S.L. (1989). Developmental consequences of prenatal exposure to methadone. Annals of the New York Academy of Sciences, 562, 195-207.

Harris, S.R. (1993). Harris Infant Neuromotor Test (Developmental Edition). Unpublished test.

Harris, S.R. (1995). [Early identification of developmental disabilities]. Unpublished raw data.

Heffelfinger, A., Craft, S., & Shyken, J. (1997). Visual attention in children with prenatal cocaine exposure. Journal of the International Neuropsychological Society, 3, 237-245.

Holzman, C., Paneth, N., Little, R., & Pinto-Martin, J. (1995). Perinatal brain injury in premature infants born to mothers using alcohol in pregnancy. Pediatrics, 95(1), 66-73.

Howard, J., Beckwith, L., Espinosa, M. & Tyler, R. (1995). Development of infants born to cocaine-abusing women: Biologic/maternal influences. Neurotoxicology and Teratology, 17(4), 403-411.

Howell, D.C. (1997). Statistical methods for psychology (4th ed.). Belmont, CA: Duxbury Press.

Hurt, H., Brodsky, N.L., Betancourt, L., Braitman, L.E., Malmud, E., & Giannetta, J. (1995). Cocaine-exposed children: Follow-up through 30 months. Developmental and Behavioral Pediatrics, 16(1), 29-35.

Jacobson, J.L., Jacobson, S.W., Sokol, R.J., Martier, S.S., Ager, J.W., & Kaplan-Estrin, M.G. (1993). Teratogenic effects of alcohol on infant development. Alcoholism: Clinical and Experimental Research,17(1), 174-183.

Johnson, J.M., Seikel, J.A., Madison, C.L., Foose, S.M., & Rinard, K.D. (1997). Standardized test performance of children with a history of prenatal exposure to multiple drugs/cocaine. Journal of Communication Disorders, 30(1), 45-73.

Koseck, K. (1999). Review and evaluation of the psychometric properties of the revised Bayley Scales of Infant Development. Pediatric Physical Therapy,11(4), 198-204.

Korner, A.F., Stevenson, D.K., Kraemer, H.C., Spiker, D., Scott, D.T., Contantinou, J., & Dimiceli, S. (1993). Prediction of the development of low birth weight preterm infants by a new neonatal medical index. Journal of Developmental and Behavioral Pediatrics, 14(2), 106-111.

Landry, S.H., Denson, S.E., & Swank, P.R. (1997). Effects of medical risk and socioeconomic status on the rate of change in cognitive and social development for low birth weight children. Journal of Clinical and Experimental Neuropsychology, 19(2), 261-274.

Levy-Shiff, R., Dimitrovsky, L., Shulman, S., & Har-Even, D. (1998). Cognitive appraisals, coping strategies, and support resources as correlates of parenting and infant development. Developmental Psychology, 34(6), 1417-1427.

Lewis, M. & Bendersky, M. (1989). Cognitive and motor differences among low birth weight infants: Impact of intraventricular hemorrhage, medical risk and social class. Pediatrics, 83(2), 187-192.

Lindenberg, C.S., & Keith, A.B. (1993). Opiate abuse in pregnancy. Annual Review of Nursing Research, 11, 249-279.

Lindsey, J.C., & Brouwers, P. (1999). Intrapolation and extrapolation of age-equivalent scores for the Bayley II: A comparison of two methods of estimation. Clinical Neuropharmacology, 22(1), 44-53.

Lipkin, P.H., & Altshuler, L.A. (1994). Early outcome determination of low-birth-weight infants using the neurodevelopmental risk examination. Clinical Pediatrics, 33(4), 398-403.

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 Sceener and the Clinical Adaptive Test/Clinical Linguistic Auditory Milestone Scale. Developmental and Behavioral Pediatrics, 19(3), 155-161.

Mattia, F.R. & deGegnier (1998). Chronic physiologic instability is associated with neurodevelopmental morbidity at one and two years in extremely premature infants. Pediatrics, 102(3), E35.

Matula, K., & Aylward, G.P. (1997). Response to commentary. Bayley scales-II. Journal of Developmental and Behavioral Pediatrics, 18(2), 112-113.

Mayes, S.D. (1997). Potential scoring problems using the Bayley Scales of Infant Development-II Mental Scale. Journal of Early Intervention,21(1), 36-44.

Medoff-Cooper, B., & Gennaro, S. (1996). The correlation of sucking behaviors and Bayley Scales of Infant Development at six months of age in VLBW infants. Nursing Research, 45(5), 291-296.

Mellins, C.A., Levenson, Jr., R.L., Zawadzki, R., Kairam, R., & Weston, M. (1994). Effects of pediatric HIV infection and prenatal drug exposure on mental and psychomotor development. Journal of Pediatric Psychology, 19(5), 617-628.

Minde, K., Goldberg, S., Perrotta, M., Washington, J., Lojkasek, M., Corter, C., & Parker, K. (1989). Continuities and discontinuities in the development of 64 very small premature infants to 4 years of age. Journal of Child Psychology and Psychiatry, 30(3), 391-403.

Nellis, L. & Gridley, B.E. (1994). Review of the Bayley scales of infant development-second edition. Journal of School Psychology, 32(2), 201-209.

O'Connor, M.J., Brill, N.J., & Sigman, M. (1986). Alcohol use in primiparous women older than 30 years of age: Relation to infant development. Pediatrics, 78(3), 444-450.

O'Connor, M.J., Sigman, M., & Kasari, C. (1993). Interactional model for the association among maternal alcohol use, mother-infant interaction, and infant cognitive development. Infant Behavior and Development,16, 177-192.

Oehler, J.M., Thompson, R.J., Goldstein, R.F., Gustafson, K.E., & Brazy, J.E. (1996). Behavioral characteristics of very-low-birth-weight infants of varying biologic risk at 6,15,and 24 months of age. Journal of Obstetric, Gynecologic and Neonatal Nursing, 25(3), 233-239.

Park, I., & Schutz, R.W. (1999). "Quick and easy" formulae for approximating statistical power in repeated measure ANOVA. Measurement in Physical Education and Exercise Science, 3(4), 249-270.

Piper, M.C., Darrah, J., Pinnell, L., Watt, M.J., & Byrne, P. (1991). The consistency of sequential examinations in the early detection of neurological dysfunction. Physical and Occupational Therapy in Pediatrics, 11(3), 27-44.

Resnick, M.B. Armstrong, S., & Carter, R.L. (1988). Developmental intervention program for high-risk premature infants: Effects on development and parent-infant interactions. Journal of Developmental and Behavioral Pediatrics, 9(2), 73-78.

Richardson, G.A., Day, N.L., & Goldschmidt, L. (1995). Prenatal alcohol, marijuana and tobacco use: Infant mental and motor development. Neurotoxicology and Teratology, 17(4), 479-487.

Robinson, B.F. & Mervis, C.A. (1996). Extrapolated raw scores for the second edition of the Bayley scales of infant development. American Journal of Mental Retardation, 100(6), 666-671.

Rosen, T.T., & Johnson, H.L. (1982). Children of methadone-maintained mothers: Follow-up to 18 months of age. The Journal of Pediatrics, 101(2), 192-196.

Ross, G., Boatright, S., Auld, P.A.M., & Nass, R. (1996). Specific cognitive abilities in 2-year-old children with subependymal and mild intraventricular hemorrhage. Brain and Cognition, 32, 1-13.

Ross, G., & Lawson, K. (1997). Commentary. Using the Bayley-II: Unresolved issues in assessing the development of prematurely born children. Journal of Developmental and Behavioral Pediatrics, 18(2), 109-111.

Ross, G., Tesman, J., Auld, P.A.M., & Nass, R. (1992). Effects of subependymal and mild intraventricular lesions on visual attention and memory in premature infants. Developmental Psychology, 28(6), 1067-1074.

Saudino, K.J., Dale P.S., Oliver, B., Petrill, S.A., Richardson, V., Rutter, M., Simonoff, E., Stevenson, J., & Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of 2-year-olds. British Journal of Developmental Psychology,16, 349-363.

Seagull, F.N., Mowery, J.L., Simpson, P.M., Robinson, T.R., Martier, S.S., Sokol, R.J., & McCarver-May, D.G. (1996). Maternal assessment of infant development: Associations with alcohol and drug use in pregnancy. Clinical Pediatrics, 35(12), 621-628.

SPSS (1999). SPSS Base 9.0 Applications Guide. Chicago, IL: Author

Streissguth, A.P., Barr, H.M., Martin, D.C., & Herman, C.S. (1980). Effects of maternal alcohol, nicotine, and caffeine use during pregnancy on infant mental and motor development at eight months. Alcoholism: Clinical and Experimental Research, 4(2), 152-164.

Tasbihasazan, R., Nettelbeck, T., & Kirby, N. (1997). Increasing the Mental Developmental Index in Australian children: A comparative study of the two versions of the Bayley Mental Scale. Australian Psychologist, 32(2), 120-125.

Thelen, E. (1995). Motor development: A new synthesis. American Psychologist, 50(2), 79-95.

Thompson, B., Wasserman, J.D., & Matula, K. (1996). The factor structure of the Behavior Rating Scale of the Bayley Scales of Infant Development-II. Educational and Psychological Measurement, 56(3), 460-474.

Thompson, R.J., Goldstein, R.F., Oehler, J.M., Gustafson, K.E., Catlett, A.T., & Brazy, J.E. (1994). Developmental outcome of very low birth weight infants as a function of biological risk and psychosocial risk. Journal of Developmental and Behavioral Pediatrics, 15(4), 232-238.

Thompson, R.J., Oehler, J.M., Catlett, A.T., & Johndrow, D.A. (1993). Maternal psychological adjustment to the birth of an infant weighing 1,500 grams or less. Infant Behavior and Development,16, 471-485.

Todd, D.A., Jana, A., & John, E. (1997). Chronic oxygen dependency in infants born at 24-32 weeks' gestation: The role of antenatal and neonatal factors. Journal of Pediatric Child Health, 33, 402-407.

Washington, K., Scott, D.T., Johnson, K.A., Wendel, S., & Hay, A.E. (1998). The Bayley Scales of Infant Development-II and children with developmental delays: A clinical perspective. Journal of Developmental and Behavioral Pediatrics, 19(5), 346-349.

Wildin, S.R., Anderson, A., Woodside, M., Swank, P., Smith, K., Denson, S., & Landry, S. (1995). Prediction of 12-month neurodevelopmental outcome from a 6-month neurologic examination in premature infants. Clinical Pediatrics, 34(6), 290-298.

Wilson, G.S. (1989). Clinical studies of infants and children exposed prenatally to heroin. Annals of the New York Academy of Sciences, 562,183-194.

Youngblut, J.M., Loveland-Cherry, C.J., & Horan, M. (1991). Maternal employment effects on family and preterm infants at three months. Nursing Research, 40(5), 272-275.

Youngblut, J.M., Loveland-Cherry, C.J., & Horan, M. (1993). Maternal employment, family functioning, and preterm infant development at 9 and 12 months. Research in Nursing and Health, 16, 33-43.

Zuckerman, B., & Bresnahan, K. (1991). Developmental and behavioral consequences of prenatal drug and alcohol exposure. Pediatric Clinics of North America, 38(6), 1387-1407.

Appendix A

Review of Studies Investigating Developmental Outcomes of Infants Exposed to Alcohol and/or Other Drugs Using the Original Bayley Scales of Infant Development (1969)

Study	Sample/ Groups	Assess-ment Points	Identification of Developmental Delays	Changes in Score Over Time	Biologic Variables	Environmental Variables
Golden et al. (1982)	Groups: 1. alcohol (n=12) 2. control (n=12)	between 6 and 20 mo.	-all group means within normal limits - alcohol-exposed group had lower MDI and PDI means than control	N/A	N/A	N/A
Jacobson et al. (1993)	Black, inner-city, alcohol-exposed (N=382)	between 11.4 and 18.5 mo.	- all group means within normal limits	- age-related decline in performance	- MDI scores inversely related to smoking and opiate use, and child's age at testing -PDI scores inversely related to parity, smoking during pregnancy. - maternal drinking during pregnancy inversely predictive of MDI and PDI scores after controlling for potential confounding variables (i.e. significant correlational relationships)	- MDI scores inversely related to maternal age and, directly related to number of prenatal visits and HOME inventory - PDI scores inversely related maternal depression and, directly related to number of prenatal visits and HOME inventory

Study	Sample/ Groups	Assess-ment Points	Identification of Developmental Delays	Changes in Score Over Time	Biologic Variable Measured	Environmental Variables Measured
O'Connor et al. (1993)	Groups: 1. alcohol (n=17) 2. control (n=17)	12 mo. Mental Scale only	- all group means within normal or accelerated range - control group had higher MDI than alcohol-exposed group	N/A	-independent variables including obstetrical complications, duration of nursing, infant gender and race, maternal age and IQ and, paternal education and social class entered first in MLR analysis (i.e. controlling for potential confounders. Alcohol use during pregnancy did not make a significant unique contribution in the prediction of MDI scores.	
O'Connor et al. (1986)	N=25 divided into abstinent to light, light to moderate, and, moderate to heavy alcohol consumption	12 mo. Mental Scale only	- all group means within normal limits	N/A	N/A	- alcohol intake prior to pregnancy inversely related to MDI scores - no relationship between maternal intelligence or education and MDI scores
Gusella & Fried (1984)	N=84 mothers who were social drinkers	13 mo.	- means not reported	N/A	-birthweight directly related to MDI and PDI scores -social drinking related to lower MDI scores	- MDI scores directly related to paternal education -pre-pregnancy drinking not related to test scores
Streissguth et al. (1980)	N=462 from larger sample of 1529 pregnant mothers to maximize number of heavy drinkers.	8 mo.	- means within normal limits - increased alcohol use related to decrease in MDI and PDI scores	N/A	- gestational age (GA) was the strongest positive predictor of MDI and PDI scores - alcohol accounted for significant variance in test scores after adjusting for parity, GA, maternal education, and intervening variables after birth	

Study	Sample/ Groups	Assess-ment Points	Identification of Developmental Delays	Changes in Score Over Time	Biologic Variables	Environmental Variables
Richardson et al. (1995)	N=737 infants exposed to alcohol / marijuana	8 and 18 mo.	- all means within normal limits, or accelerated - when sample was divided into groups by amount of substance use, there were no group differences in MDI or PDI scores	- mean MDI dropped from accelerated range at 8 mo. to normal range at 18 mo. - trends not discussed	- increase in 8 mo. MDI scores predicted by lower infant's age at examination, marijuana use and, higher work/school status - increase in 8 mo. PDI scores predicted by lower infant's age at examination, less hospitalizations, being female, lower maternal depression and, higher birthweight (BW) and work/school status - increase in 18 mo. MDI predicted by lower age at examination, stressful life events, number of siblings, higher developmental stimulation and income, and being female and White (vs. Black) - increase in 18 mo. PDI scores predicted by lower age at examination, number of people in the household, and current infant illnesses and, higher social support	
Fried & Watkinson (1988)	N=217 at 1st and N=153 at 2nd testing -only 10 heavy alcohol and 17 heavy marijuana users	12 and 24 mo.	- all means within normal limits, or accelerated	- no change in test scores over time	- increase in 12 mo. MDI scores predicted by lower parity and maternal nicotine -at 24 mo., alcohol exposure contributed predicted lower MDI scores - alcohol exposure did not contribute to PDI scores -marijuana exposure not predictive of test scores	- at 24 mo. higher HOME inventory predicted higher MDI scores

Study	Sample/ Groups	Assess-ment Points	Identification of Developmental Delays	Changes in Score Over Time	Biologic Variables	Environmental Variables
Seagull et al. (1996)	N=120 alcohol-exposed infants	12 mo.	- means within normal limits	N/A	- alcohol exposure, gender and their interaction associated with MDI scores (i.e. increased exposure and being female led to lower scores - neither alcohol consumption during pregnancy nor infant demographic variables associated with difference in PDI scores.	- maternal marital status, income, education, and number of children not associated with test scores
Chasnoff et al. (1986)	Groups: 1. opiates (n=36) 2. non-opiates (n=22) 3. control (n=34)	3, 6, 12, and 24 mo.	- all means within normal limits - isolated instances of exposed group means being lower than control	- all groups, including control, demonstrated a decrease in test scores over time	N/A	N/A
Mellins et al. (1994)	Groups: 1. HIV infected (n=24) 2. seroreverters (n=30) 3. drug exposed (n=23)	mean age of 17 mo.	- drug-exposed group MDI mean in mildly delayed range. - PDI mean within normal limits	N/A	- age negatively related to MDI, not PDI, scores - no difference in performance with respect to gender - both drug exposure and neurological functioning predictive of MDI scores - neurological functioning predicted PDI scores	- no difference in performance with respect to ethnicity

Study	Sample/ Groups	Assess-ment Points	Identification of Developmental Delays	Changes in Score Over Time	Biologic Variables	Environmental Variables
Wilson (1989)	Groups: 1. heroin (n=27) 2. methadone (n=33) 3. control (n=54)	6, 18, and 24 mo. Mental Scale only	- all means within normal limits except heroin-exposed group at 24 mos. was in mildly delayed range - heroin-exposed had lower MDI compared to control at 18 mo.	- decline in MDI scores over time for all groups, including control	N/A	N/A
Hans (1989)	Groups: 1. methadone (n=36) 2. control (n=43)	24 mo.	- all means within normal limits	N/A	- combination of methadone exposure and low SES led to lower MDI and PDI scores	
Rosen & Johnson (1982)	Groups: 1. methadone (n=41) 2. control (n=23)	6, 12, and 18 mo.	- all means within normal limits -methadone-exposed lower MDI and PDI scores at 12 and 18 mo.	- difference between methadone and control groups increased over time, especially PDI scores	- obstetrical complications, GA, BW, Apgar scores, gender, and group used as independent variables in MLR; methadone group predicted lower 18 mo. MDI and PDI scores -males obtained lower MDI scores	N/A
Howard et al. (1995)	N=74 cocaine and polydrug exposed infants	6 mo. Mental Scale only	-all means within normal limits	N/A	- higher BW predicted higher MDI scores - addiction severity and gestational age not related to MDI scores	- higher maternal sensitivity predicted higher MDI scores - HOME inventory not related to MDI scores

Study	Sample/ Group	Assess-ment points	Identification of Developmental Delays	Changes in Scores Over Time	Biologic Variables	Environmental Variables
Hurt et al. (1995)	Groups: 1. cocaine (n=101) 2. control (n=118)	6, 12, 18, 24, and 30 mo.	- all mean MDI and PDI scores within normal limits - exposed group obtained lower 18 mo. MDI scores compared to control	- MDI and PDI scores decreased over time for both groups, including control	N/A	N/A
Chasnoff et al. (1992)	Groups: 1. cocaine/ polydrug (n=106) 2. marijuana/ alcohol (n=45) 3. control (n=81)	3, 6, 12, 18, and 24 mo.	-all means within normal limits, but higher incidence of infants with scores in delayed range in exposed groups -both exposed groups had lower MDI and PDI scores at 6 mos. compared to control group - marijuana/alcohol exposed group had lower 12 mo. PDI scores than control, and lower 18 mo. PDI scores than control and cocaine/ polydrug exposed groups	- no trends reported	N/A	N/A
Billman et al. (1996)	Groups: 1. cocaine (n=46) 2. control: also high risk (n=60)	between 4 and 30 mos.	- all means within normal limits - Black cocaine-exposed had higher PDI scores than Black control	N/A	N/A	- Black infants had higher PDI scores than White infants
Johnson et al. (1997)	Groups: 1. cocaine and multiple drugs (n=14) 2. control (n=14)	between 2 and 30 mos. Mental Scale only	-drug-exposed group mean within normal limits - control group mean in accelerated performance range - control group had higher MDI scores than exposed.	N/A	N/A	N/A

Appendix B

Review of Studies Investigating Developmental Outcomes of Infants Born Prematurely Using the Original Bayley Scales of Infant Development

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Lewis & Bendersky (1989)	Groups: 1. Grade III & IV IVH (n=19) 2. Grade I & II IVH (n=13)	15 mo.	- Grade I & II IVH group means within normal limits -Grade III & IV IVH group means in mildly delayed range - infants with severe IVH had lower MDI and PDI scores compared to mild IVH	N/A	N/A	N/A
Bendersky & Lewis (1994)	N=175	between 18 and 24 mo.	- all means within normal limits	N/A	- increased number of medical complications and increase in IVH severity were predictive of lower MDI and PDI scores.	- increased family/ environmental risk and its interaction with IVH severity predicted lower MDI scores. - environmental risks did not predict PDI scores beyond biologic risk factors.
Gross et al. (1992)	Groups: 1. preterm (n=124) 2. full term (n=124)	6, 15, and 24 mo.	- all group means within normal limits. - full term group obtained higher MDI scores throughout first two years of life.	- no trends in scores reported.	- for preterm group, no differences in scores by gestational age, gender, or maternal education. - for full term group, maternal education related to MDI scores at all ages, becoming more significant with increasing age.	

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Ross et al. (1992)	Groups: 1. preterm with IVH (n=30) 2. preterm without IVH (n=30) 3. full term (n=30)	10 mo.	- all group means within normal limits. - preterm with IVH had lower MDI scores than other two groups. - no difference in PDI scores between the groups.	N/A	- gender, BW, GA, days on ventilator, duration of hospital stay, and Apgar scores were not related to MDI or PDI score	- social class, ethnic group, SES did not relate to MDI or PDI scores
Ross et al. (1996)	Groups: 1. preterm with IVH (n=27) 2. preterm without IVH (n=28) 3. full term (n=27)	10 and 24 mo.	-all group means within normal limits. - no difference in MDI or PDI scores between the three groups at either assessment point.	- no trends in scores reported.	N/A	N/A
Anderson et al. (1996)	Groups: 1. high-risk preterm (n=58) 2. low-risk preterm (n=96) 3. full term (n=119)	6 and 12 mo.	- all group means within normal limits, but more infants in high-risk group fell into moderate to severely delayed range ($p < .01$). - at 6 mo., high-risk preterm group had lower MDI and PDI scores than low-risk, who had lower scores than full term. - at 12 mo., high-risk preterm lower scores than full term.	- high risk preterm had delayed performance compared to full term infants for up to one year of age. - low risk preterm infants appeared to catch up to full term performance by one year of age.	N/A	N/A

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Wildin et al. (1995)	Groups: 1. high-risk preterm (n=49) 2. low-risk preterm (n=82) 3. full term (n=113)	6 and 12 mo.	- all group means within normal limits	- decline in MDI and PDI scores over time for all three groups, including full term infants	N/A	N/A
R.F Goldstein et al. (1995)	N=158 preterm infants with birthweight ≤1500g	6 and 24 mo.	- means not reported	- no trends reported	-at 6 mo., metabolic and respiratory acidosis inversely related to test scores, whereas at 12 mos. only metabolic acidosis was related -metabolic acidosis and hypotension predicted lower 24 mo. MDI and PDI scores	N/A
Brazy et al. (1993)	N=199 preterm infants with birthweight ≤1500g	6, 15, and 24 mo.	- means within normal limits	- no trends reported	- Neurobiologic Risk Scale (NBRS) was strongest predictor of MDI and PDI scores at all assessment points - gender was second strongest predictor of all test scores except for 24 mo. MDI score (females higher scores than males). - GA contributed to prediction of 6 mo. PDI scores	- race (White vs. Black) predicted higher 15 and 24 mo. MDI scores - higher maternal IQ led to higher 24 mo. MDI scores - at 24 mo., higher maternal education predicted higher PDI scores

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Thompson et al. (1994)	N=102 preterm infants with birthweight (BW) ≤1500g	6, 15, and 24 mo.	- all means within normal limits. - when sample was divided into high and low biological risk groups based on NRBS scores, the high risk group obtained lower MDI and PDI scores at all points - when sample was divided into psychosocial risk groups based on daily maternal stress, the high risk group obtained lower 15 and 24 mo. MDI scores	- MDI and PDI scores declined over time - number of children classified as developmentally delayed increased over time	- NBRS predicted MDI and PDI scores at all assessment points - gender predicted all test scores except 24 mo. MDI scores (i.e. females higher scores than males)	- race predicted 15 and 24 mo. MDI scores - maternal stress was a predictor of MDI scores at all points. - higher SES led to higher 24 mo. MDI scores - higher maternal education led to higher 24 mo. PDI scores
Brazy et al. (1991)	N=68 preterm infants with BW ≤1500g	6, 15, and 24 mo.	- all means within normal limits	- no trends reported	- NBRS scores predictive of MDI and/or PDI scores of less than 85. - BW, GA and, head circumference were not predictive.	N/A
Oehler et al. (1996)	N=102 preterm infants with BW ≤ 1500g	6, 15, and 24mo.	- means not reported	-categorized infants into no, continuous, and late delay. Infants with a continuous delay were less skilled motorically through 24 mo.	- Infants categorized as having a continuous delay were predominantly also of higher risk.	
Thompson et al. (1993)	Groups: 1. high distress (n= 42) 2. low distress (n=48)	6 mo.	- all means within normal limits - infants of mothers with high distress obtained lower MDI and PDI scores than infants of mothers with low distress	N/A	N/A	N/A

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Medoff-Cooper & Gennaro (1996)	Groups: 1. normal suck (n=14) 2. abnormal suck (n=5)	6 mo. Psycho-motor Scale only	- means within normal limits - abnormal sucking group obtained lower PDI scores	N/A	N/A	N/A
Cooper & Sandler (1997)	Groups: 1. 1000 to 1499g with mechanical ventilation (n=49) 2. 1000 to 1499g no ventilation (n=25) 3. < 1000g (n=25)	between 12 and 18 mo.	- group 2 and 3 means within normal limits - group 1 means in mildly delayed range - no significant difference in group means	N/A	- porencephaly strongest predictor of lower test scores -inadequate intrauterine growth and more days on supplemental oxygen led to lower test scores - infant weight, length and head circumference after discharge were not predictive of test scores.	- higher maternal education led to higher test scores. - other unspecified sociodemographic variables not predictive.
Korner et al. (1993)	N=512 Groups: 1. ≤1500g 2. 1501 to 2000g 3. 2001 to 2500 g	12 and 24 mo.	- most means within normal limits except for subgroups with the lowest BW and most severe medical complications whose means were in the mildly delayed range.	- medical complications measured by Neonatal Medical Index (NMI). - sharpest drop in scores over time for infants with most severe NMI scores.	- NMI scores mostly predictive of PDI and MDI scores for infants with BW ≤ 1500g. -NMI predictive of 24 mo. PDI scores for all groups	- higher maternal education led to higher 24 mo. MDI scores for all groups - increased maternal education less consistently predictive of PDI scores - Black or Hispanic race led to lower 24 mo. MDI scores

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Lipkin & Altshuler (1994)	N=92 preterm infants with BW <2000g	mean age of 13.7 mo. Mental Scale only	- mean MDI within normal limits - 26 infants with MDI scores < 85	N/A	Neurodevelopmental Risk Examination (NRE) score at term inversely related to MDI scores.	N/A
Feingold (1994)	N=30 preterm infants with BW <2500g	between 12 and 24 mo. Mental Scale only	- mean within normal limits - 4 infants with MDI score <84	N/A	- BW not related to MDI scores.	- maternal education and depressive symptoms not related to MDI scores - higher HOME inventory scores led to higher MDI scores
Gennaro & Stringer (1991)	N=63 preterm infants with BW <2500g	6 mo.	- MDI and PDI means within normal limits	N/A	- no relationship between infant weight, length and head circumference and MDI or PDI scores	- no relationship between maternal anxiety and depression and test scores.
Youngblut et al. (1991)	Groups: 1. employed mother (n=40) 2. non-employed (n=55) 3. leave of absence (n=15)	3 mo.	- PDI means within normal limits - MDI means of groups 2 and 3 in mildly delayed range - no significant difference in MDI or PDI scores between the groups	N/A	N/A	- employment related variables were not predictors of MDI scores. - an increase in number of hours worked predicted higher PDI scores after controlling for BW and GA

Study	Sample/ Group	Assess-ment Points	Identification of Developmental Delays	Change in Scores Over Time	Biologic Variables	Environmental Variables
Resnick et al. (1988)	Groups: 1. treatment (n=21) 2. contrast (n=20)	6 and 12 mo.	-all means within normal limits or accelerated - treatment group obtained higher MDI score at 12 mo.	- drop in PDI scores over time for both groups - drop in MDI scores over time for contrast group	N/A	N/A
Youngblut et al. (1993)	Groups: 1. employed mothers (n=37) 2. non-employed (n=29)	3, 9, and 12 mo.	- all means within normal limits - no difference in group scores at any point	- infants of non-employed mothers had an increase in MDI scores from 3 to 9 mos.	N/A	- employment related variables not predictive of MDI scores after controlling for GA. - choice in employment predictive of 3 mo. PDI scores - consistency between employment attitude and behavior predictive of 9 and 12 mo. PDI scores.
Landry et al. (1997)	Groups: 1. high-risk preterm (n=37) 2. low-risk preterm (n=42) 3. full term (n=49)	6, 12, and 24 mo. Mental Scale only	- use of mental age versus MDI standardized scores to report results - at 6 mos. high-risk group scored lower than other 2 groups	- high risk group deceleration in rate of cognitive growth, especially from 12 to 24 mo. - from 6 to 12 mo. high-risk and low-risk demonstrated less of an increase in test scores than full term group - from 12 to 24 mo. low-risk and full term groups obtained comparable increase in mental age	- higher degree of medical risk a factor in slower rates of increase in cognitive development beyond prematurity and SES	
Brady et al. (1992)	Groups: 1. SGA (n=19) 2. AGA (n=69)	4 and 12 mo.	-no significant difference between group scores at any point	- no trends in scores reported	- size for GA did not relate to significant biologic risk.	N/A

Appendix C

Descriptive Statistics for BSID-II Index Scores (Including Extrapolated Scores) For Each Group

	SHHCC				BCCH			
Descriptive	MDIex 1	PDIex 1	MDIex 2	PDIex 2	MDIex 1	PDIex 1	MDIex 2	PDIex 2
Number of scores (N)	34	34	30	30	19	19	17	19
Mean	85.71	83.71	77.17	72.53	84.05	72.79	72.71	76.16
Median	88	84	81	78.5	88	82	81	95
Standard Deviation	11.27	15.75	14.63	14.65	18.79	27.11	32.71	38.27
Variance	126.95	247.97	214.07	214.53	352.94	734.84	1069.97	1464.92
Skewness	-1.951	-.612	-1.798	-1.523	-3.080	-.987	-.782	-1.034
Std. Error of Skewness	.403	.403	.427	.427	.524	.524	.550	.524
Kurtosis	4.893	1.234	4.719	3.659	10.823	.314	-.389	-.820
Std. Error of Kurtosis	.788	.788	.833	.833	1.014	1.014	1.063	1.104
Range	44-101	37-111	24-95	23-97	15-98	15-105	15-115	15-115

Note: BSID-II mean = 100 and standard deviation = 15

MDIex 1 = Mental Developmental Index scores including scores< 50 at time 1; MDIex 2 = Mental Developmental Index scores including scores< 50 at time 2; PDIex 1 = Psychomotor Developmental Index scores including scores< 50 at time 1; PDIex 2 = Psychomotor Developmental Index scores including scores< 50 at time 2.

Appendix D

Descriptive Statistics for BSID-II Index Scores (Non-extrapolated) For Each Group

	SHHCC				BCCH			
Descriptive	MDI 1	PDI 1	MDI 2	PDI 2	MDI 1	PDI 1	MDI 2	PDI 2
Number of scores (N)	33	33	29	28	18	16	13	14
Mean	86.97	85.12	79.00	75.36	87.89	82.00	88.00	98.00
Median	88	85	81	79	89	83.5	88	99
Standard Deviation	8.65	13.62	10.83	9.95	8.81	16.69	16.90	8.72
Variance	74.91	185.48	117.29	98.98	77.63	278.67	285.67	76.00
Skewness	-1.195	.032	-.616	-.224	-1.537	-.432	.063	-.122
Std. Error of Skewness	.409	.409	.434	.441	.536	.564	.616	.597
Kurtosis	1.387	-.219	-.789	.142	3.244	-.567	-1.113	1.299
Std. Error of Kurtosis	.798	.798	.845	.858	1.038	1.091	1.191	1.154
Range	65-101	55-111	59-95	52-97	62-98	53-105	63-115	79-115

Note: BSID-II mean = 100 and standard deviation = 15

MDI 1 = Mental Developmental Index scores $\geq$ 50 at time 1; MDI 2 = Mental Developmental Index scores $\geq$ 50 at time 2; PDI 1 = Psychomotor Developmental Index scores $\geq$ 50 at time 1; PDI 2 = Psychomotor Developmental Index scores $\geq$ 50 at time 2.