Use of the Bayley Scales of Infant Development-III by Therapists for Assessing Development and for Recommending Treatment for Infants in a NICU Follow-up Clinic

June 6, 2013

This research, submitted by Katherine Lee, has been approved and accepted in partial fulfillment of the requirements for the degree of Master of Science in Occupational Therapy from the University of Puget Sound

Kari J. Tanta, PhD, OTR/L, FAOTA

Committee Chairperson: Kari J. Tanta, PhD, OTR/L, FAOTA

George S. Tomlin

Reader: George S. Tomlin, PhD, OTR/L

Yvonne Swinth, PhD, OTR/L, FAOTA

Director, Occupational Therapy Program: Yvonne Swinth, PhD, OTR/L, FAOTA

Dean of Graduate Studies: Sunil Kukreja, PhD

Abstract

Infants who have been hospitalized in a neonatal intensive care unit (NICU) may present with a multitude of challenges that put them at risk for delayed development. Early Intervention and specialized NICU follow up clinics are in place to help identify NICU graduates' need for therapy services. Well-established, standardized assessments, such as the Bayley Scales of Infant and Toddler Development (BSID-III) are utilized by occupational and physical therapists when making recommendations for therapy. The purpose of this retrospective chart review (N=104) was to identify the extent to which BSID-III motor scores were predictive of a referral for further developmental therapy in infants who were seen in NICU follow-up and to examine how therapist clinical judgment related to BSID-III scores. Independent sample t-tests conducted to compare motor performance to recommendations for motor therapy found there was a significant difference in the gross motor scores for those who were and were not recommended for motor therapy. Quality, quantity, and variability of motor skills emerged as recurring themes in therapist's clinical judgment for initiating motor therapy, despite BSID-III scores that were within normal limits. Findings from this study indicate that the factors that influence follow-up recommendations are complex and that test scores alone were not indicative of whether or not a referral was given. Information gathered from this study may help increase understanding of how BSID-III scores and clinical judgment relate for therapists recommending motor therapy for NICU graduates.

Use of the Bayley Scales of Infant Development-III by therapists for assessing development and recommending treatment for infants in a NICU follow-up clinic

Infants born extremely prematurely or with neonatal illnesses have a greater chance of surviving with the advanced technology, medical treatments, and specialized care that are now available. Neonatal intensive care units (NICUs) provide specialized care for infants who are critically ill or premature. In the past NICUs were often given a classification of level I through III based upon the sophistication of care available although more recently level IV has been proposed for use (Committee on Fetus and Newborn, 2012). There is not currently a standard classification system used by all hospitals or all states (Committee on Fetus and Newborn, 2012). A level III or IV NICU, depending on classification system used, offers the highest level of care according to this classification system and is able to provide the specialized level of care needed for infants at highest risk (Committee on Fetus and Newborn, 2012). A 2012 retrospective cohort study of 1,328,132 infants born prematurely found that mortality rates for infants born in high-level NICUs were significantly lower than those born in other lower-level delivery hospitals (Lorch, Biacci, Ahlberg, & Small, 2012). A 100 to 300 percent improvement in risk-adjusted mortality rates was seen for high-level NICUs meaning that hospitals without a high-level NICU had increased infant mortality rates (Lorch et al., 2012).

The Centers for Disease Control and Prevention found that in 2006, 77.3% of infants born with a very low birth weight were admitted into a NICU (CDC, 2010). This has led to an increase in the number of infants in need of extended specialized NICU care. These fragile infants can present with a multitude of challenges that put them at risk for delayed development (Tanta & Youngblood Langton, 2010a). Multidisciplinary teams of professionals with expertise in neonatal care work with infants and their families to develop a plan of care for high-risk newborns to promote survival. A NICU stay and neonatal complications put many infants at risk for delays in motor development, necessitating regular developmental screening through a neonatal follow-up program.

Occupational therapists and physical therapists who have substantial experience in pediatrics, advanced knowledge of development, of the medical conditions frequently seen in neonatal care, and of how to provide specialized interventions using advanced clinical reasoning, are part of the team of professionals who provide specialty care in the NICU and in follow-up settings (American Occupational Therapy Association [AOTA], 2006). Occupational therapists in the NICU setting provide feeding and neuromuscular interventions, as well as serve as a primary source for parent training and education. The AOTA (2006) declared the appropriateness of OT working in this specialty area:

> Occupational therapy's domain of concern encompassing the interaction among the biological, developmental, and social-emotional aspects of human function as expressed in daily activities and occupations makes it particularly suited to address the needs of the developing infant and family (AOTA, 2002). The occupational therapy method of activity analysis and adaptation to achieve a functional outcome is valuable in promoting "goodness of fit", as there is often a mismatch between the NICU environment, parental expectations, and the infant's capabilities (pp. 659-660).

There is much that remains unknown about the long-term developmental outcomes of infants who receive treatment in the NICU, but there is a consensus that many of these infants require continuing specialized treatment after being discharged (McGrath, Sullivan, Lester, & Oh, 2000). Much of the previous research on infants in a NICU has focused primarily on premature and very low birth weight infants, with fewer studies on those born full-term (Claas et al., 2011; Hack & Fanaroff, 1998; Huang et al., 2012; Shiariti et al., 2008). A meta-analysis of forty-one published English language studies on infants born prematurely found that there were lasting motor impairments later in childhood (Kieviet, Piek, Aarnoudse-Moens, & Oosterlaan, 2009).

Additional studies have reported sensory, cognitive, and psychosocial impairments as well (Claas et al., 2011; Hack & Fanaroff, 1999; Halsey, Collin, & Anderson, 1993; Huang et al., 2012; Kieviet et al., 2009; Stanton, McGee, & Silva, 1991; Sun, Mohay, & O'Callaghan, 2009). Studies of normal birth weight infants who were cared for in a NICU for reasons other than prematurity, have found that they too have ongoing health concerns (Marino et al., 2012, Shiariti, et al., 2008; Swanson & Dicianno, 2010). Infants admitted to a NICU may have complications resulting from congenital or genetic conditions (e.g., Down syndrome, congenital heart disease, spina bifida), from a difficult labor or delivery resulting in injury, or from illness after birth.

Despite the greater proportion of infants who need NICU care surviving there are increased risks for morbidity and poorer developmental outcomes compared with infants who do not require hospitalization following birth (McGrath et al., 2000). With screening and early recognition of developmental delay in premature infants, early interventions are possible (Rydz, Shevell, Majnemer, & Oskoui, 2005). Neonatal follow-up programs provide important services for infants who are at high risk for developmental problems after they progress home from the NICU. A survey of 170 NICU's in the U.S. found that the Bayley Scales of Infant and Toddler Development III (BSID-III) was one of the assessments used most often in follow-up programs (Kuppala, Tabangin, Haberman, Steichen, & Yolton, 2012). In addition to its clinical use the BSID-III has been widely used in research.

Gross motor development is an area where early observable delay in infants can be recognized (Spittle, Orton, Doyle, & Boyd, 2009). Gross motor skill involves control of large muscle groups that are involved in such tasks as sitting upright, walking, or moving from one position to another. As motor delays are early and visible signs of developmental concerns in infants, assessments that are able to reliably aid in identifying motor difficulties are essential for professionals who are responsible for follow-up of infants at increased risk for developmental delays.

The BSID-III includes a motor scale, which measures both fine and gross motor skills. The BSID-III is complex in its administration and interpretation thus training and experience is needed with both administration and interpretation of the assessment. A training DVD is included with the BSID-III kit and training workshops are available although not required for test administration. A competent examiner needs to possess the skills to follow standardized protocol for administration, have knowledge of statistics to understand the psychometric properties of the assessment, and be able to score and interpret the assessment (Bayley, 2006a). Occupational therapists are among the professionals who can be trained to administer this assessment, and take into account the results during clinical reasoning over follow-up treatment recommendations.

Background

In addition to infants who are born prematurely, there are a significant number of full-term infants who are admitted into the NICU. One study in Canada reported that 32% of NICU admissions were for infants born at term (Schiariti et al., 2008). Although there are most certainly differences between these groups of infants, both are at risk for increased delay in motor, psychosocial, and mental health development (Schiariti et al., 2008; Spittle et al., 2009). Infants born prematurely are at risk for developmental delays due to a multitude of factors associated with premature birth. Premature birth puts infants at risk for respiratory distress, anemia, intraventricular hemorrhage, neuromotor problems such as cerebral palsy, visual impairments, hearing impairments, learning difficulties, and psychosocial behavioral problems (March of Dimes, 2012).

Infants born at term may also be hospitalized in a NICU due to complications from a difficult delivery, respiratory problems, birth defects, or other congenital diseases that require specialized care (Schiariti et al., 2008). Because of the fragile state of infants in the NICU extended NICU care may be needed. A 2007 study found that the average length of stay in the NICU for 502 infants born prematurely was 2.88 weeks (Berns, Boyle, Popper, & Gooding, 2007). The cost of neonatal intensive care is high, both monetarily and for the toll it places on family members, but it has been shown to increase survival in infants and result in better outcomes later in life (Wilson-Costello, 2007).

Infants who graduate from the NICU leave with a unique set of medical problems. The American Academy of Pediatrics affirmed the importance of follow-up programs for pre-term infants at high-risk for developmental delay (American Academy of Pediatrics, 2008). Federal law such as the Individuals with Disabilities Education Act mandates early identification and intervention services for children with developmental disabilities (Mulligan, 2003; Palfrey, 2009). NICU follow-up programs are utilized by many hospitals with a NICU to facilitate continuity of clinical care and make any necessary referrals to early intervention programs for further follow-up care. A survey of 194 NICUs affiliated with pediatric residency programs found that 93 percent reported affiliation with a follow-up program (Kuppala et al., 2011).

Most NICU follow-up programs are multidisciplinary with neonatologists and other professionals such as occupational therapists and physical therapists working together to provide multidisciplinary care for the infants (Kuppala et al., 2011). NICU follow-up clinics are a source of reassurance and support for families with the expertise available from therapists. They help ensure that appropriate diagnoses are made, refer to needed services including occupational therapy and physical therapy, and offer assistance with the coordination of care. Infants who are at risk for developmental delay and who do not attend a follow-up clinic have been reported to have higher incidence of motor disabilities, lower cognitive skills, and less access to early intervention when compared with similar infants who do attend a follow-up clinic (Callahan et al., 2001; Campbell et al., 1993; Slater, Naqvi, Andrew, & Haynes, 1987; Tin, Fritz, Wariyar, & Hey, 1998; Wolke, Sohne, Ohrt, & Riegel, 1994). This means that when there is a hindrance in care there is greater cost for infants, their families, and the healthcare system (Catlett, Thompson, Johndrow, & Boshkoff, 1993). Therefore, if reliable measures for predicting which infants and families would benefit most from early intervention programs could be established the cost-effectiveness of care would increase.

Critical development in infancy. Periods of critical development occur during the early stages of life and early identification of infants at risk for developmental delays can help ensure that infants receive appropriate interventions. Critical development occurs in areas such as play skills, self-help skills, and oral-motor and feeding skills, primitive reflex patterns, development of automatic reactions, development of fine and gross motor skills, and cognitive development (Mulligan, 2003).

Premature infants often lack the motor control and central nervous system maturity that would enable them to move into a flexed and midline position independently. Because of this infants born prematurely may not be able to achieve movements and positioning as their full term peers would. The ability to experience a flexed and midline position is important in that it facilitates hand-to-mouth activity, promotes flexor tone development, helps prevent deformities of a positional origin, and promotes a calm state (Hunter, 2005). Early in infancy movements are largely reflex based but before the first year of life is over most of those primitive type reflexes are integrated into more complex and voluntary movements (Mulligan, 2003).

Because of the rapid changes that occur in development at these early stages, accurately assessing development and predicting the need for further treatment can be challenging. The overarching goal of early identification of children with developmental delays is to obtain follow-up services for those in need through programs designed to maximize potential development. Assessment tools that are comprehensive, have strong psychometric properties, and are cost-effective and easy to administer are essential to identifying children in need of follow-up services. It is also important to determine whether a test administered in infancy can predict developmental functioning at later points in time. Predictive validity aids those caring for infants in a NICU and follow-up clinic in making decisions about the need for early intervention in infants who are at risk of developmental disabilities.

Importance of appropriate screening assessments. The ability of a test to measure what it claims to measure otherwise known as accuracy of a test, is often established through sensitivity and specificity. Tests that lack sensitivity may miss identifying infants who need services and tests that are overly sensitive may unnecessarily refer infants to services that are not necessarily needed. A test with high specificity would be able, most of the time, to correctly identify infants who do not have developmental delay. Longitudinal assessments of outcomes for infants who received neonatal intensive care can aid those who work with them in understanding the implications of the care received and to evaluate the effectiveness of interventions. Standardized assessments of development are also important for evaluating treatment outcomes and determining eligibility for early intervention programs. In a prospective longitudinal study of infants with a very low birth weight (VLBW), such infants were found to use 2.8 times more special academic assistance than infants born with a normal birth weight in grades kindergarten through sixth grade. (Schraeder, Heverly, O'Brien, & Goodman, 1997). Additional studies on VLBW infants found a similar need for academic assistance later in life (Hack, Klein, & Taylor, 1995; Lindeke, Stanley, Else, & Mills, 2002; Saigal, Szatmari, Rosenbaum, Campbell, & King, 1991).

BSID-III

History of the Bayley Scales of Infant Development. The Bayley scales have been used extensively for over four decades in both clinical assessments and research to identify infants and toddlers with developmental delay and to provide information that will help inform intervention planning (Bayley, 2006a). The test was first published in 1969 by Nancy Bayley as *The Bayley Scales of Infant Development* and was designed for children from one month to forty-two months of age. Twenty-four years later a second version of the test, The Bayley Scales of Infant Development 2nd edition (BSID-II) was published. The BSID-II revised normed populations to include data from a wider range of groups including those born prematurely, prenatally exposed to drugs, deprived of oxygen during birth, those with developmental delay, autism, and Down Syndrome (Bayley, 2006a). The age range was extended with the second edition and psychometric properties were strengthened (Bayley, 2006a). The BSID-II included a mental, motor and behavior rating scale. The motor scale addresses body control, gross motor, and fine manipulation skills (Bayley, 1993)

The most current version, the BSID-III, was released in 2006. The BSID-III was normed using a contemporary population of infants from 2000, making it better suited for current use in comparing infants than the previous version that was standardized in 1988 (Bayley, 1993; 2006a). Data were collected for children with commonly prescribed diagnoses and basal and ceiling levels were extended (Bayley, 2006a). The updated BSID-III has separate composite scores for motor, cognition, and language and has scale scores measuring receptive communication, expressive communication, and gross and fine motor development (Bayley, 2006a; 2006b). Subtests for social-emotional and adaptive behavior were added in addition to a screening test for further testing (Bayley, 2006a). New additions also included a scoring assistant for test administrators, and growth charts to track progress over time (Bayley, 2006b). A parent-report questionnaire has been included in the test intended to measure social-emotional and adaptive behavior (Bayley, 2006a; 2006b). Steps were taken on an overarching basis to make the test more administrator friendly and make test items more motivating for infants and children being assessed. Changes were also made to the existing motor scale to rearrange some of the fine and gross motor items in order to increase content validity and add new items (Bayley, 2006a).

Predictive value. The BSID-III is frequently used in the assessment of infant development and in research; however its validity as a predictor for need for further treatment has not been reliably established (Anderson, De Luca, Hutchinson, Roberts, & Doyle, 2010). A 2010 descriptive prospective cohort study conducted in Australia comparing the BSID-II to the updated BSID-III found that developmental delay was underestimated in Australian children using the 3rd edition, however, their study was limited to children at 2 years of age and the Social-Emotional and Adaptive Behavior scales of the test were not administered (Anderson et al., 2010). The appropriateness of the BSID-III and its value as a discriminatory and predictive tool for children at other age groups has yet to be widely studied in the U.S. or elsewhere.

BSID-III in research. Recent research on the sensitivity of the BSID-III to identify when there is developmental delay has been inconclusive. Findings have suggested that the cognitive composite scale of the test and the BSID-III in general, may be overestimating ability in infants

(Anderson et al., 2010; Vohr et al., 2012). Anderson et al. (2010) suggested that the small number of published studies using the BSID-III indicating enthusiasm for the BSID-III may have declined in response to reports that it may be overestimating development. In their 2010 prospective cohort study of extremely low birth-weight (ELBW) Australian two-year-olds who were administered the BSID-III, they concluded that the test underestimated delay in those children. Subjects were 211 infants born weighing less than 2.2 pounds or before 28 weeks gestation who survived to age 2. A control group of 202 infants born at 37 weeks gestation or later and weighting over 4.3 pounds was used. The means for the control group were also higher than expected with composite scores that ranged from 0.55 to 1.23 SD above the normed mean for the test (Anderson et al., 2010). The rates of developmental delay found in these ELBW 2-year-olds was below previously reported rates found using similar age bands (Anderson et al., 2010). Proportions of those with motor, language, and cognitive delay were found to be 16%, 21%, and 13% respectively (Anderson et al., 2012). There were limitations to this study, however, including that no description of exactly where within the broader categories delays were found was given, the Social-Emotional and Adaptive Behavior scales of the test were not administered, and data were collected for a single age group. Generalizability to countries outside of Australia may also be limited.

Motor development in infants has been well researched to create reliable norms for comparison. A retrospective study of 93 infants with a history of NICU stay found that the Gross Motor Scale of the BSID-III was able to identify those infants eligible for follow-up services in early intervention (Jackson, Needleman, Roberts, Willet, & McMorris, 2012). Infants in the study were between 6 and 8 months corrected age (Jackson et al., 2012). All but 5 of the children accepted into early intervention were categorized as being in the “at risk” range or the "emergent" range (Jackson et al., 2012). Regression analysis was used comparing the BSID-III to the Alberta Infant Motor Scale in predicting acceptance into early intervention. The Gross Motor Scale was found to account for a significant amount of the variance in early intervention service acceptance (Jackson et al., 2012). Five of those accepted to early intervention were scored as "competent" on the Gross Motor Scale of the test which could be an indicator that the test did not accurately identify these infants' needs for services (Jackson et al., 2012). Follow-up of this subgroup showed that medical needs were the qualifying factor for admission to early intervention services (Jackson et al., 2012).

Currently, the BSID-III is widely used as a tool in outpatient NICU practices including early intervention and specialized follow-up clinics (Kuppala et al., 2012). Establishing a better understanding of development for sub-groups of infants with a history of NICU stay using Bayley scores would allow professionals working with them to better predict future performance levels and identify which infants to refer for follow up services.

Therefore, the purpose of this study was to (1) identify the extent to which BSID-III motor scores are predictive of a need for further motor therapy in infants who were seen in an NICU, in order to best develop follow-up protocols and plans of care for treatment in neonatal follow up clinics, and (2) examine how clinical judgment relates to BSID-III motor scores when therapists make recommendations for further motor therapy.

Method

Research Design

This study was a retrospective chart review of infants who were evaluated between January 2011 and September 2012 in one Pacific Northwest hospital-based NICU follow-up clinic, in order to determine the predictive accuracy of BSID-III scores for later needed motor therapy as determined by occupational therapists and physical therapists who conducted the evaluation sessions. For the purpose of this paper the term “motor therapy” was chosen to describe therapy services that would be recommended for infants for whom gross motor and/or fine motor skill concerns arose. In a NICU follow-up clinic such as the setting for the present study, motor therapy services infants were referred for were carried out by either an occupational therapist or physical therapist depending primarily on availability of therapists as either discipline would be equally qualified to provide services to infants at this age. Percentages of infants who received multiple BSID-III tests at different time periods were calculated. The study compared the BSID-III motor scores for infants seen in a NICU follow-up clinic at first and subsequent assessments where the Bayley was administered with therapist recommendations for follow-up or referral to begin motor therapies.

Prior to the chart review, an occupational therapist and a physical therapist met with the researchers to explore the information needs of therapists when making recommendations about the need for motor therapy. This was done in order to ensure that data collected were relevant, thereby increasing the likelihood of findings from the study being of value to therapists. Gross motor (GM), Fine Motor (FM) and Motor Composite scores for infants who were administered the BSID-III were gathered and compared to the clinical judgments made by NICU follow-up therapists when making a recommendation for further motor therapy.

Setting

The NICU Follow-up Clinic where the study was conducted provides comprehensive multidisciplinary evaluations for infants who were hospitalized in its level III neonatal intensive care unit and received services from the Children’s Therapy NICU Team. Primary care providers may refer other infants, including those not hospitalized in the NICU, for assessments in the program. Physical or occupational therapists, and speech language pathologists assess gross motor, fine motor, cognitive, expressive language, receptive language and feeding skills. If indicated, a dietician and a nurse practitioner may assess some infants.

There are two primary and two back-up therapists who administer the BSID-III in the NICU follow-up clinic. Infants are typically seen at 4 months adjusted age and then at 3-6 month intervals depending on their needs. Infants are followed through 2 years of age, or until they have been able to achieve motor, cognitive, language and feeding skills appropriate for their chronological ages. If indicated, families are provided with activities and exercises to help support skill development, and are occasionally referred for outpatient therapy services.

Participants

The population of interest for this study was all infants seen in a NICU follow-up clinic between specified dates when there was consistency in the type of medical charting system in use to make data extraction simpler and more reliable. A list of infants meeting the following criteria was obtained through the hospital's information technology department: (1) seen at the NICU follow-up clinic between January 2011 and September 2012 for an initial or follow-up evaluation; (2) billed for occupational therapy or physical therapy under a NICU follow-up evaluation code; and (3) infant was less than one year old at the time of service.

A total of 364 infants met the criteria during a search of charts. Of those 364 infants a convenience sample of 213 infants with last names beginning with letters "A" through "M" was used by researchers for the purpose of obtaining a large enough N to conduct desired statistical analyses and due to time constraints. Fifty-seven percent or 122 infants met all inclusion criteria, which in addition to those described above also required a history of NICU hospitalization be found upon examination of charts. Reasons for excluding some infants who were on the initial list included: they were only seen by speech language pathology (SLP) when evaluated in the NICU follow-up clinic, no previous history of NICU hospitalization was found, no NICU follow-up evaluation found, and a small number of charts reviewed were found to be incomplete with missing test scores. BSID-III motor scores were collected when available. There were a total of 104 infants for which complete BSID-III motor scores were available. Participant medical records were accessed through the hospital's digital records repository system, Chartmaxx, for review. Some infants had evaluations completed at several age increments while others did not.

Instrumentation

A review of medical records was conducted through a digital records repository system in order to gather pertinent data from charts of infants who met inclusion criteria. BSID-III raw scores for the GM and FM subtests as well as motor composite scores were collected. Other descriptive and demographic information that was collected from the chart review for data analysis included gender, length of NICU stay, birth weight, degree of prematurity, gestational age at birth, adjusted age at evaluation, actual age at evaluation, overall impressions from therapists, observed plagiocephaly, observed head turn preference, and recommendations for further treatment. Descriptive information regarding overall impressions from therapists following NICU follow-up clinic evaluations was also recorded to add to the discussion and to be used for later research. Information gathered was used in order to develop a descriptive profile of BSID-III scores and demographics for the sample.

The BSID-III has established psychometric properties. The average internal consistency reliability coefficient by age for the BSID-III motor composite scale is .92 (Bayley, 2006a) Internal consistency of the motor composite scale of the test ranged from .72 to .95 (Bayley, 2006a). Test-retest reliability for the motor scale ranged from .79 to .84. Concurrent validity and construct validity have all been established for this assessment. Concurrent validity with the Peabody Developmental Motor Scales-Second Edition (PDMS-2) for the Motor Composite Scale was .57 and .59 for fine and gross-motor subtests respectively (Bayley, 2006b). The construct validity between the Motor Composite Scale and GM and FM subtests was found to be .71 and .69 (Bayley, 2006b).

Therapists at the clinic where the study took place had not completed any assessments of their own inter-rater reliability. Through speaking with the program director it was determined that group discussions and agreement were often made between experienced occupational therapists and physical therapists in day-to-day discussions; however, it was not believed that the therapists differ significantly in either their test administration or interpretation (K. Tanta, personal communication, October, 2012).

Procedures

Institutional review board approval was first obtained from the university. Approval from the Research Oversight Committee at the hospital was then granted. An occupational therapist and physical therapist from the hospital where the study took place who both had experience in administering the BSID-III were consulted throughout this study. Following university and hospital approval the researcher was trained in how to access and extract desired data from computerized therapy charts. The charts for infants relevant to the study were made accessible by the information technology department of the hospital. The program lead for the Children’s Therapy department at the hospital then trained the researcher on how to navigate the online charting system. Consensus was made among committee members on how to arrange the Excel spreadsheet and for all coding that was implemented.

Charts from all infants seen in the NICU follow-up clinic from January 2011 to September 2012, who met inclusion criteria described above, were reviewed. A pilot phase of data extraction occurred. Ten charts were reviewed after which a discussion among the research team occurred and decisions were made to include the aforementioned variables as well as the addition of age equivalences for both GM and FM subtests as projected by the BSID-III. Alterations made to the original data collection form were submitted to the university IRB. Data were collected onto an Excel spreadsheet. Steps were taken to eliminate any unnecessary identifying information. The list of possible subjects with identifying information was kept in a locked file cabinet at the hospital.

Incidence of cases where an infant never underwent standardized assessment using the BSID-III when evaluated in the NICU follow-up clinic was collected to calculate frequency of occurrence. Data were grouped by infant for ease in comparison of multiple BSID-III administrations. Inter-rater agreement for data collection was completed to ensure that all researchers were able to collect data in a consistent method. Review of digitized paper medical charts took place in the hospital in a private room. Data did not leave the hospital until after the removal of identifying information.

BSID-III scores were collected from both initial and follow-up evaluations found in participant charts. BSID-III GM raw scores, FM raw scores, motor composite scores, GM age equivalents, FM age equivalents, percentiles and standard deviations were collected and recorded into the data file for analysis. Age measurements reported in months and weeks in charts were converted into days following data collection in order to establish consistency among data for the purpose of performing statistical analyses with the same unit of measurement. Because this was a retrospective chart review no inter-rater reliability for administration of the BSID-III was known.

Numerical data were recorded onto the spreadsheet. Time periods were converted into days for statistical analysis. A numerical coding system was implemented for variables when appropriate. For example infants who were administered the motor portion of the BSID-III received a “1” whereas infants who were not were given a “2.” Degree of prematurity was classified using gestational age at birth. Infants were given a classification of very premature, moderate prematurity, mild prematurity, or full term using guidelines used by Kramer et al. (2000).

Narrative information from therapists was also extracted from charts related to observations and clinical decision-making by the primary researcher. Narrative writing from infants’ charts was read through initially while being transcribed into the data sheet during the data collection phase and once more following completion of the data set. Common themes were identified regarding reasons for recommending therapy and particular areas of concern in motor skills observed during evaluations by therapists. When reading through narrative data the primary researcher wrote down words/phrases used by therapists that were mentioned repeatedly when describing their reasoning for recommending initiation of motor therapy. These notes were then used for initial data analyses of narrative information collected.

Data Analysis

Statistical Package for the Social Sciences (SPSS) 21.0 was used to perform the data analysis. Descriptive statistics for the demographic factors were calculated for each variable. The proportion of males and females was examined for even distribution as well as any differences in birth weight, length of NICU stay, and gestational age at birth that could represent possible group differences by gender. Frequencies were calculated for the variables plagiocephaly, head turn preference, and whether or not each of eight possible boxes regarding follow-up recommendations was checked by therapists (see Figure 1). Plagiocephaly and head turn preference were separated out after consulting with therapists in the NICU follow-up clinic during which it was gathered those factors may weigh more heavily on decisions to recommend follow-up therapy.

Frequencies were calculated for how many BSID-III administrations infants received. Average length of NICU stay was calculated using Microsoft Excel 2011. A descriptive profile of BSID-III scores by first and subsequent test administrations was created using mean, standard deviation, and range for GM, FM, and motor composite scores. Groups were created to allow for further comparisons to be made. Examples of groups used for certain analyses included level of prematurity, gender, and those with or without identified plagiocephaly.

To help answer the question of whether or not BSID-III motor scores are predictive of therapists' recommending motor therapy percentage comparisons were used. Age equivalency scores for GM and FM subtests were used along with infants' adjusted age to calculate a difference score as a way of categorizing infants based on their performance on motor subtests. Henceforth age equivalence minus adjusted age will be referred to as "performance deviation" for this study. The percentage of time infants were referred for further therapy when they received an age equivalence score no further than 30 days below their adjusted age was calculated as well as the percentage of those who received a similar score and were not recommended for therapy. The same analyses were performed for infants whose performance deviation was found to be more than 30 days behind based upon age equivalency scores and adjusted age. It should be noted that the classification system used for performance deviation was derived as a way of performing initial analyses without the use of standard scores, which were not consistently reported in charts. The 30-day criterion was selected to provide a way of classifying those who performed close to or above their adjusted age from those who performed below expected per adjusted age.

An examination whether or not there was agreement between the BSID-III motor performance and therapist recommendations for therapy was made. Infants who were found to have greater than a 30 day delay per performance deviation were placed into one group (yes motor therapy) whereas those who had less than or equal to a 30 day delay were placed in a second group (no motor therapy). Contingency tables were created for both GM and FM subtests using therapist's decision of "yes" or "no" for motor therapy versus BSID-III subtest scores indicating need for motor therapy as described above. Cases were sorted into sub-groups where the BSID-III score agreed with therapist recommendations and where they did not. Percent agreement (yes/yes for therapy and no/no for therapy) was calculated, as well as non-agreement (yes/no and no/yes).

Themes and trends identified during initial review of qualitative narrative based data regarding therapists' clinical decision-making were analyzed qualitatively. Preliminary findings regarding reasons for whether or not initiation of motor therapy was thought to be necessary by evaluating therapists were reported. Words and phrases that came up when there was agreement between therapists and BSID-III scores were compared to those that were seen when there was disagreement between the two. Because there is perhaps more interest in cases where the BSID-III and therapist clinical judgment differs more time was spent reviewing cases where disagreement occurred.

Results

To obtain a target sample size of 100 viable cases 213 charts were reviewed and 91 infants were excluded due to not meeting inclusion criteria. Reasons for exclusion included lack of previous NICU hospitalization, lack of occupational therapy or physical therapy evaluation (several infants were found to have been only evaluated by speech-language pathology), and sixteen infants (13.1%) had been evaluated but never administered the BSID-III and two charts were found to be missing one or more motor subtest scores. Following review of medical records through a digital records repository 104 cases were found that met all inclusion criteria and were thus used for data analysis. Infants were administered the BSID-III by occupational therapists 26% of the time and by physical therapists 74% of the time in charts reviewed.

Description of Infants

Table 1 shows the baseline characteristics of the 104 infants who were included in this study. Infants included in analyses ranged in age from 99 (3 months, 3 days) to 470 days (15 months, 6 days) chronological age and 63 (2 months, 1 day) to 470 days (15 months, 6 days) adjusted age. There was no statistically significant difference in proportion of female (n = 50) and male infants. Furthermore, comparable gender distributions were also found for all degrees of prematurity in this study. Percentages of infants who were considered to be born full term, mildly premature, moderately premature, and very premature are presented in Table 2. The mean birth weights for infants were as follows: full term, 2892.71 grams. (SD = 813.34); mildly premature, 2324.15 grams (SD = 450.69); moderately premature, 1840.63 grams (SD = 398.56); very premature, 1230.13 grams (SD = 333.97). Diagnoses other than prematurity among infants were not recorded for the purpose of this study.

BSID-III Motor

Infants whose charts were reviewed for this study were administered the GM and FM subtests of the BSID-III between one and five times during subsequent follow-up evaluations by either an occupational therapist or physical therapist at the clinic. The range in age for the first administration of the BSID-III was from 63 days to 274 days which is a noteworthy length of time this early in life and makes generalizing test administration with particular age distributions difficult. Table 3 shows frequency of BSID-III administrations. Also presented in Table 3 are average ages at testing and average age equivalences generated from BSID-III GM and FM subtests.

Follow-up Recommendations Made by Therapists

The NICU follow-up clinic where the study was completed used a template for initial evaluations and re-evaluations of infants to help guide evaluations. Therapists were able to write-in their clinical observations and decisions in an area titled "Overall Impressions." Therapists were then able to check boxes next to standard recommendations for families such as initiation of home programs, follow-up evaluations, and initiation of motor therapy. Figure 1 shows descriptions of possible boxes that could be checked. Table 4 shows percentages of boxes checked during first and subsequent BSID-III administrations. The box most frequently checked regardless of which BSID-III administration was for "Follow-up neurodevelopmental evaluation in __weeks / months in order to monitor progress, identify concerns, and determine need for therapy services secondary to risks associated with prematurity and/or history of NICU care". Initiation of motor therapy was most likely to occur during infants' first (19.2%) or second (14.5%) time being tested with the BSID-III.

Age Equivalence Versus Adjusted Age

By subtracting infants' adjusted ages from their age equivalents as determined by the BSID-III pairwise comparisons could be made. On the first BSID-III administration where the largest sample, *N* = 104, was present, there was very little difference between adjusted age and age equivalence generated by the BSID-III for FM or GM. Infants tended to score at or slightly above their adjusted ages on the motor subtests of the BSID-III. Outliers present influenced both the mean scores calculated and standard deviations, which therefore may be overstating how variable the age difference scores were. A positive correlation existed between adjusted age and age equivalence scores for FM and GM, which was expected because as age increases one would expect motor skill development to progress as well.

Age equivalence scores for GM and FM subtests of the BSID-III were compared with infants' adjusted ages (age equivalence in days minus adjusted age in days at time of testing equals performance deviation). This was done in order to determine the degree of difference between the two and whether or not differences were positive or negative (indicating performance above expected for adjusted age) or negative (below that expected per adjusted age), respectively. In the clinic where the study was conducted adjusted ages are used until age two which is consistent with the BSID-III Administration Manual (Bayley, 2006b). Thus adjusted ages were used when performing analyses for the purpose of this study. Currently consensus among professionals regarding correcting a child's age for prematurity is lacking. A recent study examining the frequency and impact of using corrected age found corrected age to be used more frequently by primary care providers with resulting impacts on assessment and care recommended (D'Agostino, 2013).

Descriptive statistics for performance deviation for GM and FM subtests at first and subsequent BSID-III administrations are presented in Table 5 and Table 6, respectively. For all administrations of the BSID-III the means for GM and FM performance deviation were positive (meaning infant was developing ahead of expectations) with the exception of GM on the second administration of the BSID-III. When differences were examined for performance deviation separating cases by the degree of prematurity there was a higher number of means that were negative for both GM and FM. Descriptive statistics accounting for degree of prematurity are presented in Table 7 for FM and Table 8 for GM**.** The decrease in *n* with each subsequent BSID-III administration appeared at each prematurity level.

Figure 2 displays a box graph for the calculated age differences for GM and FM subtests for the first BSID-III administration (N = 104). The median for the distribution of scores for both GM and FM is near zero on the graph. The height of the inner box was slightly greater for GM than FM indicating more variability in the middle 50% of the scores. Outliers were detected in both the positive and negative direction for both subtests ($n = 6$).

Figure 3 displays the distribution of performance deviations for both GM and FM age equivalences compared to adjusted age. For FM the mean performance deviation was 2.81 with a standard deviation of 22.57. For GM the mean performance deviation was 4.08 with a standard deviation of 27.66. Both distributions of performance deviations calculated for GM and FM subtests followed a normal distribution as depicted in Figure 3. A single sample t-test was used to test the null hypothesis that the paired difference was zero, meaning the age equivalences generated from BSID-III motor subtest scores was equal to the adjusted age for infants. For GM performance deviation no statistically significant difference between age equivalence and adjusted age was found, $t(103) = 1.503$, $p = .136$. For FM performance deviation, there was no statistically significant difference between age equivalence and adjusted age, $t(103) = 1.269$, $p = .207$.

Performance Deviation and Follow-up Recommendation's By Therapists (First BSID-III Administration)

There were 20 infants (19.2%) who were recommended to begin motor therapy services by therapists following their first evaluation with the BSID-III. Because the average age of infants who were recommended for motor therapy was similar to that of those who did not receive a recommendation (133.1 versus 134.5 days respectively) raw scores were included in analysis. For GM, infants who were recommended for motor therapy had an average raw GM score of 16.0 compared to a mean of 18.4 for those not recommended for motor therapy. The mean FM raw score for infants who were recommended for motor therapy was 13.0 compared to a mean of 14.4 for those who were not. The average motor composite score for infants who were recommended for motor therapy was 98.7 compared to an average motor composite score of 105.4 for those who did not receive a recommendation.

Independent samples t-tests were used to examine the relationship between performance deviation and whether or not recommendations were made for initiation of motor therapy. Performance deviation for GM for two sub-groups (whether or not recommendations were made to initiate motor therapy) was found to be statistically significantly different, $t(102) = 2.47$, $p = .015$. When FM performance deviation for the first BSID-III administration was compared for the sub-groups whether or not initiation of motor therapy was recommended, no significant difference was found ($p = .094$).

For the second BSID-III administration the mean GM and FM scores of infants recommended for motor therapy were 25.9 (SD = 7.85) and 22.5 (SD = 5.01) respectively. For infants who were not recommended for motor therapy GM raw scores averaged 32.1 (SD = 6.88) and for FM the mean raw score was 24.1 (SD = 3.95). An independent samples t-test was run for whether or not motor therapy was recommended and GM and FM subtest raw scores. For GM, a difference was found, $t(53) = 2.33$, $p = .024$, indicating that there was a real difference between those who did and did not receive a recommendation for motor therapy (infants who scored lower being more likely to receive a referral). For FM no difference was found ($p = .31$).

(Infants who scored lower being more likely to receive a referral).

Percentage of Infants Recommended for Motor Therapy

Frequencies of motor therapy recommendations were run for infants who had performance deviations that were less than or equal to 30 days delay and for those who had a performance deviation of greater than 30 days delay per BSID-III testing for both GM and FM subtests. For infants who had a FM performance deviation representing less than or equal to 30 days delay (*n* = 97), a motor therapy recommendation was made 18.8% of the time. For infants whose FM performance deviation represented delay of greater than 30 days (*n* = 7) a motor therapy recommendation was made 28.6% of the time. For GM, when performance deviation was less than or equal to 30 days delay (*n* = 96) a motor therapy recommendation was made 17.9% of the time. When performance deviation was greater than 30 days delay for the GM subtest (*n* = 8), a motor therapy recommendation was made 37.5% of the time.

Contingency tables for GM and FM performance deviation at first BSID-III testing and whether or not a recommendation was made for initiation of motor therapy are shown in Figure 4. There was 78.8% agreement found between GM cut-score and therapist judgment and 77.9% agreement for FM. When disagreement was present between performance deviation cut-scores and whether or not therapists felt initiation of motor therapy was indicated 77% of the time (GM) and 78.8% of the time (FM) it was because the BSID-III indicated less than a 30 day delay per age equivalence and the therapist referred the infant for motor therapy. Cohen's kappa was calculated for both GM and FM subtest. For GM, K = .114 and for FM, K = .053.

Qualitative Data Related to Therapists' Clinical Judgment

Initial review of narrative data collected from charts found that therapists most often identified areas of concern related to motor skill development rather than cognitive development.

Preliminary analysis of narrative data collected found concerns related to quality, quantity, and variety of motor skills emerged as reoccurring themes in therapist's clinical judgment for initiating motor therapy even when BSID-III motor scores were reported as within normal limits (WNL). One therapist commented,

> Motor skills appropriate for age per BSID [III]. Clinical observations reveal some differences in movement with muscle recruitment and quality that are slightly concerning…These differences may negatively impact motor skill progression and will be addressed in home program and then possibly motor tx [treatment] if progress is not noted at next evaluation (Unknown Occupational Therapist, 2011).

Other reasons therapists gave for warranting initiation of motor therapy despite WNL BSID-III scores was the presence of plagiocephaly (characterized by flattening of one side of the skull), preference for using one side of the body during activities such as crawling or reaching, muscle tone abnormalities, and imbalance in preference for upper or lower extremity use.

In a small number of cases therapists noted that infants' skills demonstrated during testing were inconsistent with parent report of performance when at home. In all cases where this occurred parent reported indicated increased skill level at home compared to during testing. Instances where therapists reported that skills observed during testing were believed to be below actual skill level were not found in initial review of narrative portions of charts. Therapists did however report cases where formal testing using the BSID-III was not initiated or completed due to clinical judgment. Reasons given for not administering the BSID-III during evaluations were due to the very young age of the infant, arousal state during evaluations (ex: very tired after immunizations), and irritable behavior during testing that interfered with standardization. In one case it was reported that testing was not completed due to a father stating that the family needed to leave abruptly after reporting not understanding why the appointment was necessary due to there being nothing wrong with his child.

Individualized home programs were developed by therapists and provided to families a majority of the time even when initiation of motor therapy was not recommended. No two home programs were found to be the same between infants. Examples of recommendations given in home programs included specific handling and positioning of infants, increased time spent in prone (tummy time), encouraging reach through use of toys, promoting play in different positions (such as in prone and while seated), and facilitation from one posture to another such as from sit to stand). Through reading narrative data collected from charts therapists reported spending a significant amount of time educating family members on how to perform activities recommended and they often gave family members the opportunity to demonstrate their understanding of suggested exercises/activities during the session. Families were also provided with handouts to supplement recommendations demonstrated by therapists during the sessions and contact information to call the therapist if any questions arose.

Some infants were referred to outside clinics or for home-based services. In either case this was reported to work better with the family for reasons including not having reliable transportation to the setting clinic, sibling already receiving therapy elsewhere, and other clinics being closer in proximity to the family home. When a home-based service referral was given the therapist often commented that a home-based service was recommended after discussing options with family and concluding that that type of service would work better for them.

Discussion

The first purpose of the study was to identify the extent to which BSID-III motor scores were predictive of a need for further motor therapy in infants seen in one NICU follow-up clinic.

In the current study, 19.2% of infants received a referral for motor therapy following their first BSID-III evaluation, indicating that there was a large portion of infants seen in the clinic who did not receive a referral for subsequent motor treatment. Results indicated that, when defining a BSID-III delay as an age equivalent score more than 30 days behind the infant's adjusted age, therapist recommendations agreed with BSID-III scores about 78% of the time. When there was disagreement, it was more than three times as likely to be when the BSID-III indicated no delay, but the therapist recommended motor therapy. The GM and FM contingency tables were almost identical indicating no significant difference of GM or FM being a better predictor for whether or not motor therapy is recommended. Infants whose GM and FM performance deviations showed a delay of greater than 30 days below expected for their adjusted age were recommended for motor therapy more frequently than those who showed less than or equal to a 30 day delay. The higher percentage of infants being referred for motor therapy when performance was below average suggests that there is a correlation between performance deviation and motor therapy recommendations.

The second purpose of the study was to examine how clinical judgment related to BSID-III motor scores. Preliminary review of justifications given by therapists for whether or not to refer an infant for motor therapy revealed that the decision is complex with the therapist considering factors from motor skills observed to a family's access to transportation. Therapist concerns regarding quality, quantity, and variety of motor skills during evaluation sessions emerged as reoccurring justification for the initiation of motor therapy services despite WNL BSID-III motor scores. These findings along with findings from the contingency table discussed above suggest that therapists in this study were proactive in initiating motor therapy services despite the absence of a large motor delay. The use of individualized home programs was found for a high number of infants regardless of whether or not a recommendation for motor therapy was given. These may have been sufficient to progress infants who showed more minor motor developmental delays resulting in fewer infants receiving a motor therapy referral later on.

Heterogeneity among infants seen in NICU follow-up is often reported in the literature as a limitation to conclusions that can be drawn. Not only do infants present with differing medical and developmental histories, but factors such as the family dynamic, cultural and physical environments they are a part of, access to financial resources, access to healthcare services, and access to transportation to get to and from therapy appointments also vary. Through analysis of therapist justification given for whether or not motor therapy treatment was recommended it appeared that therapists were taking into account each infant's unique set of client factors as well as the context of their family and home environments. This was apparent when in-home therapy was recommended due to a lack of transportation options and in frequency of therapy recommended to accommodate for already busy family schedules.

Infants who received a motor therapy recommendation at the time of their first BSID-III were performing at slightly above their expected developmental level for both gross and fine motor. This means that recommendations for initiating motor therapy were still made for infants who were performing above average. This could be due to the BSID-III not capturing a full picture of infants' developmental status. Possible support for this hypothesis was found in the narrative portion of therapist's notes that were reviewed. Therapists were found to repeatedly express concern over quality, quantity, and variety of movements despite BSID-III motor scores that were WNL. Therapists also cited the presence of plagiocephaly as a reason for initiating motor therapy despite WNL motor scores. Concern over the potential for plagiocephaly to influence development is supported by a recent study that found a correlation between plagiocephaly and later developmental delay in toddlerhood (Hutchison, 2012).

Whether or not there was a difference in the ability of motor subtests of the BSID-III to predict subsequent recommendation for initiation of motor therapy was of interest to the current study team. A study of 85 infants born prematurely who were administered the BSID-III reported in the technical manual found that FM subtest scores were able to differentiate premature infants from those born at or near term (Bayley 2006a). In contrast, the current study found that there was a significant difference for GM performance deviation scores when compared to whether or not an infant was recommended motor therapy and no difference was found for FM. Findings from the current study are supported by a study conducted by Jackson et al., which found that the GM subtest of the BSID-III was able to identify infants who were later determined to be eligible for early intervention services (2012). It should be noted that Jackson et al. (2012) did not examine the FM subtest scores.

Therapists working in a NICU follow-up clinic are working with families and infants who are in the early stages of transitioning home from the NICU and therefore have the ability to greatly influence how the transition goes. Following discharge from the NICU, therapists who see these infants and their families can use their clinical observation skills to be attentive to how the family is dealing with the transition and what type of follow-up care would best fit with the family as a whole. Furthermore, these therapists may have a history of working with families while their infant was in the NICU and may therefore have established rapport that would lead them to feel comfortable sharing concerns. When families feel comfortable enough to share concerns with therapists appropriate referrals may be made so specialists will also be able to provide needed support to families.

One area where the therapist-caregiver relationship has potential for impact is on attrition rates in NICU follow-up. Attrition was likely responsible for a significant proportion of the decrease in numbers of infants seen for subsequent BSID-III testing. Although exact attrition numbers are not known for the current study the program lead of the clinic reported that there is a high level of attrition present (K. Tanta, personal communication, May 2013). Difficulties with attrition during NICU follow-up have also been reported in the literature (Ballantyne et al., 2012).

Through discussion with setting therapists, another hypothesis for the drop in number of infants seen for subsequent BSID-III testing was due to infants who were graduating (increase in scores from first to last testing observed) from follow-up developmental assessments, indicating that the clinic was functioning in an ethical manner and not continuing to treat those who were doing well. Furthermore, it was hypothesized that infants who return to the clinic for subsequent testing were likely to be from families where caregivers are very diligent and did not want to miss any opportunities to further the infants' progress. Narrative data revealed that there were also infants who were referred to other clinics for their follow-up closer to home or in-home that also contributed to some attrition observed.

Premature birth has been found to increase stress levels and incidence of depression for some caregivers (Korja et al., 2008). Studies have found that implementation of family-centered interventions, both during hospitalization and when transitioning home, have a positive impact on rates of maternal stress and depression, self-esteem, and infant-parent interactions (Meyer, 1994; Korja et al., 2012). Findings from the current study indicated that individualized and family centered home programs were utilized frequently. Therapists in the current study were also found to be spending time going over home programs with families and explaining their importance while also giving opportunities for parents to handle and practice observing and implementing recommendations with their infants directly. This may have increased the rate of feelings of competency in ability to carry out home program recommendations as well as follow through with home programs recommended and been partially responsible for the low rate of infants who were recommended for motor therapy that was found.

Implications for Occupational Therapy

Overall findings from the current study suggest that there is added value in the ability of therapists to detect subtleties in development that standardized tests alone are unable to detect. This gives some validation to the work that therapists think they are doing. Early detection of developmental delay allows for initiation of early intervention services which have been found to positively impact motor development in infants (Blauw-Hospers, 2005). In order to best detect these subtleties therapists must have well-established observation and clinical judgment skills. Both the American Occupational Therapy Association and American Physical Therapy Association have published work on the advanced training and skills needed for therapists to work with this specialized population (AOTA, 2006; Sweeney, 2009). Training programs for both disciplines need to ensure that ample opportunity is given to develop clinical judgment and observation skills. In order to allow for increased time spent developing these skills focus on test administration procedures could be limited as the number of assessments in use currently is high and growing and which tests are used also varies by setting. Professors and employers could also improve upon assessment of therapist's ability to pick up on subtleties during observation and reasoning behind clinical decision making during schooling and in the workplace.

Therapists working in NICU follow-up settings are in a unique position to help support and empower families in their transition to home life and in the care of their infant. When working with such young infants it is important to factor the entire family into development of an occupational profile. By fostering a clinic environment where families feel free to express any concerns or achievements building of rapport between therapist and caregivers can be more successful. Caregivers are experts on the infants they care for and should be made to feel that their opinions and observations are validated. One way to acknowledge the individuality of infants seen and empower families is through the development and recommendation of individualized home programs. When caregivers are given adequate guidance in how to carry out recommendations they can then be the ones administering therapy and can gain feelings of confidence in their ability to care for their infants.

Furthermore, when there is good follow through with home programs the need for later motor therapies can be decreased which is also cost-effective for families and allows for therapists time to be spent seeing infants who have greater need. In cases where follow-through with home programs is suspected to be low, therapists should re-evaluate what was asked of families and try to make recommendations that will fit within the routine of the family. Although therapist contributions to the development of infants seen in NICU follow-up is arguably significant it is caregivers who influence development each day and therapists should work to advocate and support families and their individual needs.

Limitations

As this study was conducted at a single hospital in the Pacific Northwest generalizability may be limited. There are two primary and two back-up therapists who administer the BSID-III in the clinic and their inter-rater reliability has not been formally assessed for their administration and scoring. Because of time constraints only about two thirds of the potential eligible subjects charts were reviewed. Although the number of charts eligible for review is relatively large uneven distribution of factors including level of prematurity and age at each testing were present. With each administration of the BSID-III the number of infants with motor scores decreased making any analyses completed less powerful. Using only adjusted age for analyses performed such as when calculating difference from age equivalents generated from the BSID-III may have underestimated the presence of developmental motor skills delays.

Standard deviations based upon motor composite scores were not reported consistently in charts limiting data that could be used for analyses. As mentioned previously only using adjusted age may have influenced the ability to detect significant differences for variables that were compared. Caution is advised when using age equivalents for interpretation as they are unable to provide information relative to performance of similar aged peers and the potential for small changes in raw score to disproportionately affect age equivalencies generated (Bayley 2006a). Furthermore, the “performance deviation” calculation used in analysis was not standard practice, limiting comparisons that can be made to other studies. Although the chart system accessed was computerized most documents accessed were hand written and then scanned into the system allowing for the possibility of misinterpreting information due to legibility of handwriting. A significant number of charts initially reviewed did not meet inclusion criteria (42.7%) for reasons including lack of NICU hospitalization, and incomplete or lack of BSID-III test scores due to not having been administered the assessment (evaluated only by SLP or too young for first testing).

Implications for Future Research

Further statistical analyses using existing data collected from this study is warranted. Clinical decision-making could be analyzed further through thorough analyses of narrative data collected from charts regarding therapists’ overall impressions from evaluations. Standard scores could be calculated from data already collected with additional time and would allow for comparisons to be made with similar studies. Examining the consistency of BSID-III motor scores over time would also be beneficial in helping answer the purpose of this study. Further studies looking at both GM and FM subtests as predictors for early intervention services is also warranted due to conflicting findings in current research.

Conducting a survey or qualitative study where therapists who administer the BSID-III frequently are questioned about their perceptions of agreement between their assessment of an infant's developmental skills and the test would add to the findings of this study and help guide further research as well as have potential to influence future test development. A study investigating follow through with motor therapy recommendations would also be beneficial as the percentage who actually initiated motor therapy treatment was not known in this study.

Conclusion

This study investigated the use of the Bayley Scales of Infant Development-III by therapists for assessing motor skill development in infants and their recommendations for treatment in a NICU follow-up clinic. Initial findings from this study indicate that the factors that influence whether or not an infant is recommended for follow-up therapy are complex and that a test score alone is not indicative of whether or not a referral will be given.

Upon initial analysis it appears that following initial testing therapists rely on clinical judgment in addition to BSID-III motor scores when recommending motor therapy. Quality, quantity, and variety of movements observed emerged as areas of concern frequently reported by therapists despite BSID-III scores that were WNL. NICU follow-up programs are cost intensive clinics with significant investment required by facilities that establish and maintain them, clinicians, and the families that attend them. It is therefore important for therapists to have a strong understanding that the standardized assessments they use may not fully correspond with their own observations, so that they can best identify those in need for follow-up early intervention services. Therapists who evaluate these infants are charged with the difficult task of taking into consideration those complexities and deciding what type of care will best serve the infant in the attainment of developmental skills.

References

American Academy of Pediatrics Committee on Fetus and Newborn. (2008). Policy Statement. Hospital discharge of the high-risk neonate. *Pediatrics, 122*, 1119-1126.

American Occupational Therapy Association. (2006). Specialized knowledge and skills for occupational therapy practice in the neonatal intensive care unit. *American Journal of Occupational Therapy, 60*, 559-668.

Anderson, P. J., De Luca, C. R., Hutchinson, E., Roberts, G., & Doyle, L. W. (2010). Underestimation of developmental delay by the new BSID-III Scale. *Archives of Pediatrics and Adolescent Medicine, 164*, 352-356.

Ballantyne, M., Stevens, B., Guttmann, A., Willan, A.R., & Rosenbaum, P. (2012). Transition to neonatal follow-up programs: Is attendance a problem? *Journal of Perinatal & Neonatal Nursing, 26*, 90-98.

Bayley, N. (1993). *Bayley Scales of Infant Development, (2nd ed.)*. San Antonio, TX: The Psychological Corporation.

Bayley, N. (2006a). *Bayley Scales of Infant and Toddler Development–Third edition: Technical Manual*. San Antonio, TX: Harcourt Assessment.

Bayley, N. (2006b). *Bayley Scales of Infant and Toddler Development-Third edition: Administration Manual*. San Antonio, TX: NCS Pearson.

Berns, S. D., Boyle, M. D., Popper, B., & Gooding, J. S. (2007). Results of the premature birth national need-gap study. *Journal of Perinatology, 27*, S38-S44.

Blauw-Hospers, C. & Hadders-Algra, M. (2005). A systematic review of the effects of early intervention on motor development. *Developmental Medicine and Child Neurology, 47*, 421-432.

Callahan, C., Doyle, L., Rickards, A., Kelly, E., Ford, G., & Davis, N. (2001). Children followed with difficulty: How do they differ? *Journal of Pediatrics and Developmental Health, 37*, 152-156.

Campbell, M. K., Halinda, E., Carlyle, M. J., Fox, A. M., Turner, L. A., & Chance, G. W. (1993). Factors predictive of follow-up clinic attendance and developmental outcome in a regional cohort of very low-birth-weight infants. *American Journal of Epidemiology, 138, 704*-713.

Catlett, A. T., Thompson, R. J., Johndrow, D. A., & Boshkoff, M. R. (1993). Risk status for dropping out of developmental follow-up for very low-birth-weight infants. *Public Health Reports, 108*, 589-594.

Centers for Disease Control and Prevention. (2010). *MMWR Weekly: Neonatal intensive-care unit admission of infants with very low birth weight – 19 states, 2006*. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5944a4.htm

Claas, M. J., De Vries, L. S., Bruinse, H. W., Van Haastert, I. C., Uniken Venema, M. M. A., Peelen, L. M., & Koopman, C. (2011). Neurodevelopmental outcome over time of preterm born children ≤ 750 g at birth. *Early Human Development, 87*, 183-191.

Committee on Fetus and Newborn. (2012). Levels of Neonatal Care. *Pediatrics, 130*, 587-597.

D'Agostino, J. A., Gerdes, M., Hoffman, C., Manning, M. L., Phalen, A., & Bernbaum, J. (2013). Provider use of corrected age during health supervision visits for premature infants. *Journal of Pediatric Health Care, 27*, 172-179.

Hack, M., & Fanaroff, A. (1998). Outcomes of children of extremely low birthweight and gestational age in the 1990's. *Early Human Development, 53*, 193-218.

Hack, M., Klein, N., & Taylor, G. (1995). Long-term developmental outcomes of low birth weight infants. *The Future of Children, 5,* 176-196.

Halsey, C. I., Collin, M. F., & Anderson, C. L. (1993). Extremely low birth weight children and their peers: A comparison of preschool performance. *Pediatrics, 91*, 807-811.

Huang, J. H., Huang, H. L., Chen, H. L., Lin, L. C., Tseng, H. I., & Kao, T. J. (2012). Inattention and development of toddlers born in preterm and with low birth weight. *Kaohsiung Journal of Medical Sciences, 28,* 390-396.

Hutchison, L. Deformational plagiocephaly is associated with developmental delay in toddlers. *Journal of Pediatrics, 160,* 527-528.

Jackson, B. J., Needelman, H., Roberts, H., Willet, S., & McMorris, C. (2012). Bayley Scales of Infant Development Screening Test-Gross Motor Subtest: Efficacy in determining need for services. *Pediatric Physical Therapy, 24*, 58-62.

Kieviet, J. F., Piek, J. P., Aarnoudse-Moens, C. S., & Oosterlaan, J. (2009). Motor development in very preterm and very low-birth-weight children from birth to adolescence: A meta-analysis. *Journal of the American Medical Association, 302*, 2235-2242.

Korja, R., Latva, R., & Lehtonen, L. (2012). The effects of preterm birth on mother-infant interaction and attachment during the infant's first two years. *Acta Obstetricia et Gynecologica Scandinavica, 91,* 164-173.

Korja, R., Savonlahti, E., Ahlqvist-Bjorkroth, S., Stolt, S., Haataja, L., Lapinleimu, H., Piha, J, & Lehtonen, L. Maternal depression is associated with mother-infant interaction in preterm infants. *Acta Paediatr, 97,* 724-730.

Kramer, M. S., Demissie, K., Yang, H., Platt, R. W., Sauve, R., & Liston, R. (2000). The contribution of mild and moderate preterm birth to infant mortality. *Journal of the American Medical Association, 284,* 843-849.

Kuppala, V. S., Tabangin, M., Haberman, B., Steichen, J., & Yolton, K. (2012). Current state of high-risk infant follow-up care in the United States: Results of a national survey of academic follow-up programs. *Journal of Perinatology, 32,* 293-298.

Lindeke, L., Stanley, J. R., Else, B. S., & Mills, M. M. (2002). Neonatal predictors of school-based services used by NICU graduates at school age. *American Journal of Maternal Child Nursing, 27*, 41-46.

Lorch, S. A., Baiacchi, M., Ahlberg, C., & Small, D. S. (2012). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics, 130*, 270-278.

March of Dimes (2012). Your premature baby. Retrieved from http://www.marchofdimes.com/baby/premature_indepth.html

Marino, B. S., Lipkin, P. H., Newburger, J. W., Peacock, G., Gerdes, M., Gaynor, J. W.,… Mahle, W. T. (2012). Neurodevelopmental outcomes in children with congenital heart disease: Evaluation and management: A scientific statement from the American Heart Association. *Circulation, 126*, 1143-1172.

McGrath, M. M., Sullivan, M. C., Lester, B. M., & Oh, W. (2000). Longitudinal neurologic follow-up in neonatal intensive care unit survivors with various morbidities. *Pediatrics, 106,* 1397-1405.

Meyer, E. C., Garcia Coll, C.T., Lester, B.M., Zachariah, C.F., McDonough, S.M, & Oh, W. (1994). Family-based intervention improves maternal psychological well-being and feeding interaction of preterm infants. *Pediatrics, 93*, 241-246.

Mulligan, S. (2003). *Occupational therapy evaluation for children: A pocket guide*. Philadelphia, PA: Lippincott Williams & Wilkins.

Palfrey, J. S. (2009). Legislation for the eduation of children with disabilities. In W. L. Coleman,

A. C. Crocker, H. M. Feldman, & E. R. Elias (Eds.), *Developmental-behavioural pediatrics* (4th ed.) (pp. 972-975). Philadelphia, PA: Saunders Elsevier.

Rydz, D., Shevell, M. I., Majnemer, A., & Oskoui, M. (2005). Topical review: Developmental screening. *Journal of Child Neurology, 20,* 4-21.

Saigal, S., Szatmari, P., Rosenbaum, P., Campbell, D., & King, S. (1991). Cognitive abilities and school age performance of extremely low birth weight children and matched controls at age 8 years: A regional study. *Journal of Pediatrics*, *118,* 751-760.

Schiariti, V., Klassen, A. F., Hoube, J. S., Synnes, A., Lisonkova, S. & Lee, S. K. (2008). Perinatal characteristics and parents' perspective of health status of NICU graduates born at term. *Journal of Perinatology*, *28,* 368-376.

Schmmler, C. J. & Hunter, J. G. (2005). Neonatal intensive care unit. In J. Case-Smith (Ed.), *Occupational therapy for children* (5th ed.) (pp. 688-770). St. Louis, MO: Elsevier Mosby.

Schraeder, B. D., Heverly, M. A., O'Brien, C., & Goodman, R. (1997). Academic achievement and educational resource use of very low birth weight (VLBW) survivors. *Pediatric Nursing*, *23*, 21-25.

Slater, M. A., Naqvi, M., Andrew L., & Haynes, K. (1987). Neurodevelopment of monitored versus nonmonitored very low-birth-weight infants: The importance of family influences. *Journal of Developmental Behavior Pediatrics, 8,* 278-285.

Spicer, A., Pinelli, J., Saigal, S., Wu, Y. W., Cunningham, C., & DiCenso, A. (2008). Health status and health service utilization of infants and mothers during the first year after neonatal intensive care. *Advanced Neonatal Care, 8*, 33-41.

Spittle, A., Orton, J., Doyle, L. W., & Boyd, R. (2009). Early developmental intervention programs post hospital discharge to prevent motor and cognitive impairments in preterm infants (Review). *The Cochrane Library, 1*, 1-71.

Stanton, W. R., McGee, R., & Silva, P. A. (1991). Indices of perinatal complications, family background, child rearing, and health as predictors of early cognitive and motor development. *Pediatrics, 88,* 954-959.

Sun, J., Mohay, H., & O'Callaghan, M. (2009). A comparison of executive function in very preterm and term infants at 8 months corrected age. *Early Human Development, 85*, 225-230.

Swanson, M. E., & Dicianno, B. E. (2010). Physiatrists and developmental pediatricians working together to improve outcomes in children with spina bifida. *Pediatric Clincs of North America, 57*, 973-981.

Sweeney, J.K., Heriza, C.B., & Blanchard, Y. (2009). Neonatal physical therapy. Part I: Clinical competencies and neonatal intensive care unit clinical training models. *Pediatric Physical Therapy, 21*, 296-307.

Tanta, K. J., & Youngblood Langton, S. (2010a). NICU primer for occupational therapists: Exploring the needs of fragile infants, the context in which they are cared for, and the role of OT in this specialized practice area – Part I of II. *Journal of Occupational Therapy, Schools, & Early Intervention, 3*, 179-186.

Tanta, K. J., & Youngblood Langton, S. (2010b). NICU primer for occupational therapists: Therapeutic staffing trends in northwest neonatal intensive care units – Part II of II. *Journal of Occupational Therapy, Schools, & Early Intervention, 3*, 268-281.

Tin, W., Fritz, S., Wariyar, U., & Hey, E. (1998). Outcome of very preterm birth: Children reviewed with ease at 2 years differ from those followed up with difficulty. *Archives of Disease in Childhood, 79,* 83-87.

Vohr, B. R., Stephens, B. E., Higgins, R. D., Bann, C. M., Hintz, S. R., Das, A., . . . Fuller, J. (2012). Are outcomes of extremely preterm infants improving? Impact of Bayley Assessment on outcomes. *Journal of Pediatrics, 161*, 222-231.

Wilson-Costello, D. (2007). Is there evidence that long-term outcomes have improved with intensive care? *Seminars in Fetal & Neonatal Medicine, 12,* 344-354.

Wolke, D., Ratschinski, G., Ohrt, B., & Riegel, K. (1994) The cognitive outcome of very preterm infants may be poorer than often reported: An empirical investigation of how methodological issues make a big difference. *European Journal of Pediatrics, 153,* 906-915.

Table 1

Participant Characteristics (*N* = 104)

	Mean	Minimum	Maximum	SD
	N = 104			
Gestational Age	33.8	25.9	41.7	24.46
Birth Weight	2070.4	580	4608	719.09
Length of Stay	104	5	149	24.46

Note. Abbreviations: GA, gestational age in weeks; BW, birth weight in grams; Length of stay, number of days in the neonatal intensive care unit.

Table 2

Degree of Prematurity for Participants (*N* = 104)

Classification	*n* (%)
Full Term	14 (13.5)
Mildly Premature	48 (46.2)
Moderately Premature	19 (18.3)
Very Premature	23 (22.1)

Note. Classifications are based on gestational age at birth. Definitions: Full term, 37 weeks and greater; Mildly premature, 34-36 weeks; Moderately Premature, 32-33 weeks; Very premature, less than 32 weeks.

Table 3

Age and BSID-III Scores for Subsequent Test Administrations

BSID-III	Age		BSID-III Motor Scores			Age Equivalence	
Administration (*N*)	Chronological	Adjusted	FM	GM	Composite	FM	GM
1 (104)	5 m 9 d	4 m 4 d	14.16	17.93	104.07	4 m 5 d	4 m 6 d
2 (55)	10 m 1 d	8 m 5 d	23.87	31.22	98.96	8 m 7 d	8 m 2 d
3 (25)	12 m 9 d	11 m 5 d	28.24	40.56	102.04	11 m 7 d	11 m 6 d
4 (10)	16 m 7 d	15 m 1 d	32.30	47.40	104.89	16 m 3 d	15 m 2 d
5 (1)	18 m 9 d	15 m 20 d	33.00	49.00	103.00	17 m 0 d	16 m 0 d

Note. Abbreviations: BSID-III, Bayley Scales of Infant and Toddler Development, Third Edition; FM, fine motor, scores are raw; GM, gross motor, scores displayed are raw.

BSID-III age equivalences are based on adjusted age.

Table 4

Percent of Recommendation Boxes Checked

	% Checked Per BSID III Administration (*N*)				
Recommendation	1	2	3	4	5
Floor Time	72.1	56.4 (31)	36	10	0
Home Program	75	67.3 (37)	52	50	100
Motor Therapy	19.2	14.5 (8)	4	10	0
Orthotics	1	1.8 (1)	0	0	0
ND Eval to Monitor Progress	81.7	90.9 (50)	84	60	100
Follow-up Evaluative Screens at 4-6 Years					
Age	41.3	65.5 (36)	88	70	0
ND Eval at 4 Months	0	0 (0)	0	0	0
Other	38.5	34.5 (19)	24	80	0
N	104	55	25	10	1

Note. Recommendations described in more detail in Figure 1.

Table 5

Descriptive Statistics for Gross Motor Performance Deviation: Age Equivalent - Adjusted Age

BSID III Administration (*N*)	GM Age Equivalent - Adjusted Age				
	Mean	Range	Minimum	Maximum	SD
1 (104)	4.1	241	-124	117	27.66
2 (55)	-9.7	201	-106	95	38.26
3 (25)	3.9	133	-72	61	31.41
4 (10)	10.7	113	-58	55	37.14
5 (1)	10	0	10	10	0

Note. Abbreviations: GM, gross motor; SD, standard deviation.

Measurements in days.

Table 6

Descriptive Statistics for Fine Motor Performance Deviation: Age Equivalent - Adjusted Age

BSID III Administration (*N*)	FM Age Equivalent - Adjusted Age				
	Mean	Range	Minimum	Maximum	SD
1 (104)	2.8	181	-94	87	22.57
2 (55)	2	204	-106	98	48.92
3 (25)	6.3	334	-124	210	61.92
4 (10)	34.7	243	-86	157	96.83
5 (1)	40	0	40	40	0

Note. Abbreviations: FM, fine motor; SD, standard deviation.

Measurements in days.

Table 7

Descriptive Statistics: Gross Motor Performance Deviation by Prematurity

Degree Premature	BSID III Administration (*n*)	GM Age Equivalent - Adjusted Age				
		Mean	Range	Minimum	Maximum	SD
Full Term	1 (14)	5.4	68	-35	33	21.65
	2 (9)	2.3	59	-28	31	20.81
	3 (6)	6	78	-24	54	32.18
	4 (2)	48	96	41	55	9.9
	5 (0)					
Mildly Premature	1 (48)	8.1	214	-97	117	28.58
	2 (23)	**-4.7**	143	-70	73	31.67
	3 (13)	5.1	133	-72	61	34.74
	4 (6)	10.8	66	-29	37	30.88
	5 (0)					

	1 (19)	1.7	83	-40	43	23.12
	2 (9)	**-14**	88	-53	35	28.6
Moderately Premature	3 (3)	9	61	-30	31	33.87
	4 (0)					
	5 (0)					
	1 (23)	**-3.2**	160	-124	36	32.02
	2 (14)	**-22.7**	201	-106	95	57.21
Very Premature	3 (3)	**-10.3**	40	-36	4	22.28
	4 (2)	**-27**	62	-58	4	43.84
	5 (1)	10	0	10	10	0

Table 8

Descriptive Statistics: Fine Motor Performance Deviation by Prematurity

Degree Premature	BSID III Administration (*n*)	FM Age Equivalent - Adjusted Age				
		Mean	Range	Minimum	Maximum	SD
Full Term	1 (14)	6.8	66	-35	31	17.65
	2 (9)	40.1	115	-17	98	36.52
	3 (6)	-4	82	-54	28	32.91
	4 (2)	48	46	25	71	32.53
	5 (0)					
Mildly Premature	1 (48)	7	122	-35	87	22.77
	2 (23)	1.1	132	-59	73	41.25
	3 (13)	7.4	200	-124	76	50.21
	4 (6)	60.8	229	-72	157	109.88
	5 (0)					

	1 (19)	**-8.3**	72	-35	37	19.29
	2 (9)	**-9.6**	113	-73	40	37.83
Moderately Premature	3 (3)	79	209	1	210	114.14
	4 (0)					
	5 (0)					
	1 (23)	0.7	128	-94	34	25.08
	2 (14)	**-13.4**	190	-106	84	63.29
Very Premature	3 (3)	**-50.3**	87	-86	1	45.57
	4 (2)	**-57**	58	-86	-28	43.84
	5 (1)	40	0	40	40	0

Sample of Recommendations boxes for Follow-up

☐ Initiation of home program activities targeting:________________________________

☐ Initiation of home program activities targeting:________________________________

☐ Implementation of motor therapy services in order to address the above-stated concerns. Goals will be established by primary therapist.

☐ Orthotics:____

☐ Follow-up neurodevelopmental evaluation in ______ weeks / months in order to monitor progress, identify concerns, and determine need for therapy services secondary to risks associated with prematurity and/or history of NICU care.

☐ Follow-up evaluative screens for sensory processing, motor coordination, and language/communication at 4-6 years of age, prior to beginning school. These evaluations are recommended secondary to risks associated with prematurity that may interfere with learning and classroom readiness.

☐ Follow-up neurodevelopmental evaluation at 4 months adjusted age with completion of testing with standardized evaluation in order to monitor progress, identify concerns, and determine need for therapy services secondary to risks associated w/ prematurity and/or history of NICU care.

☐ Other:____

Figure 1. Recommendation Boxes for Follow-up

Boxplot: Performance deviation (age equivalent minus adjusted age) for fine and gross motor subtests

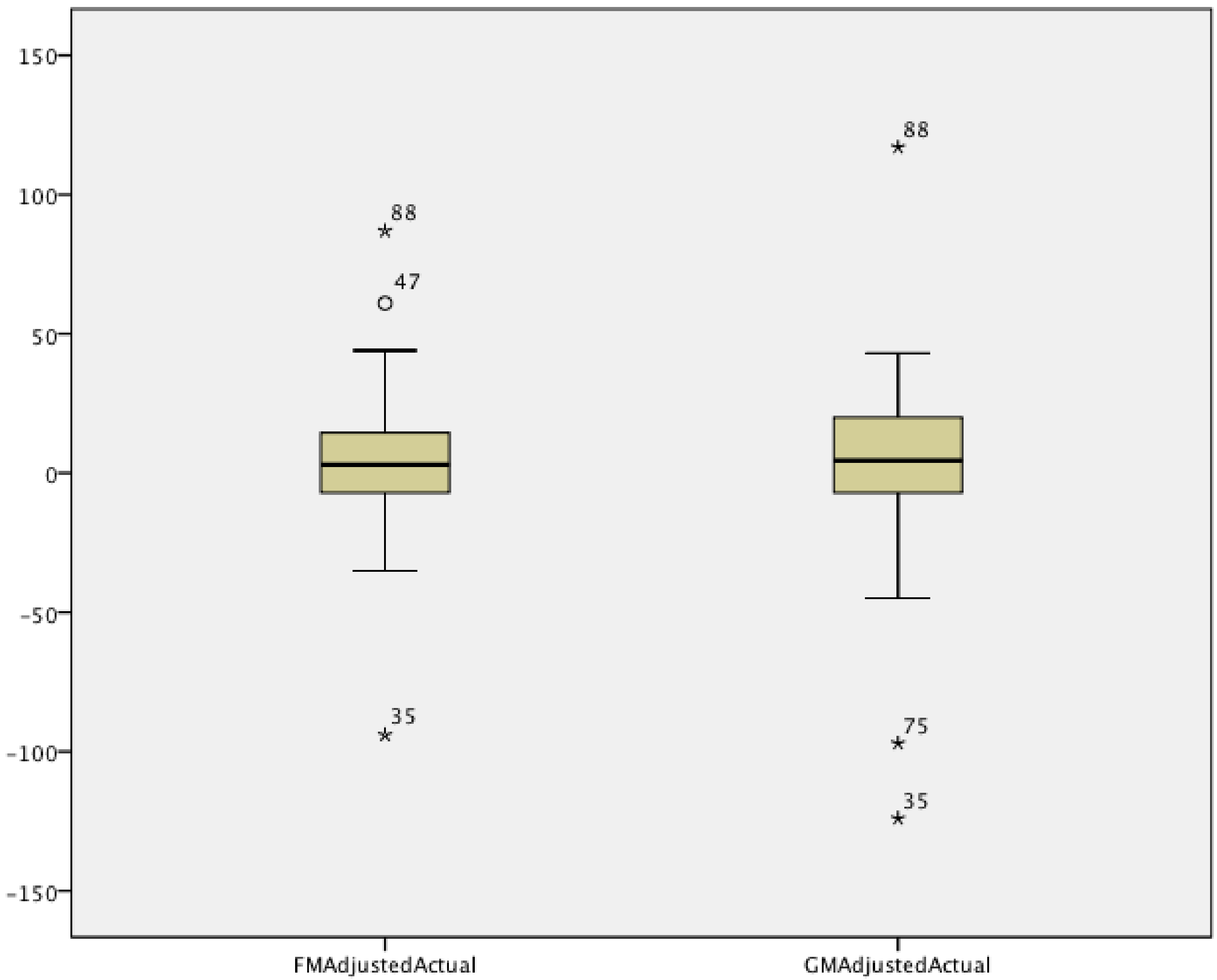
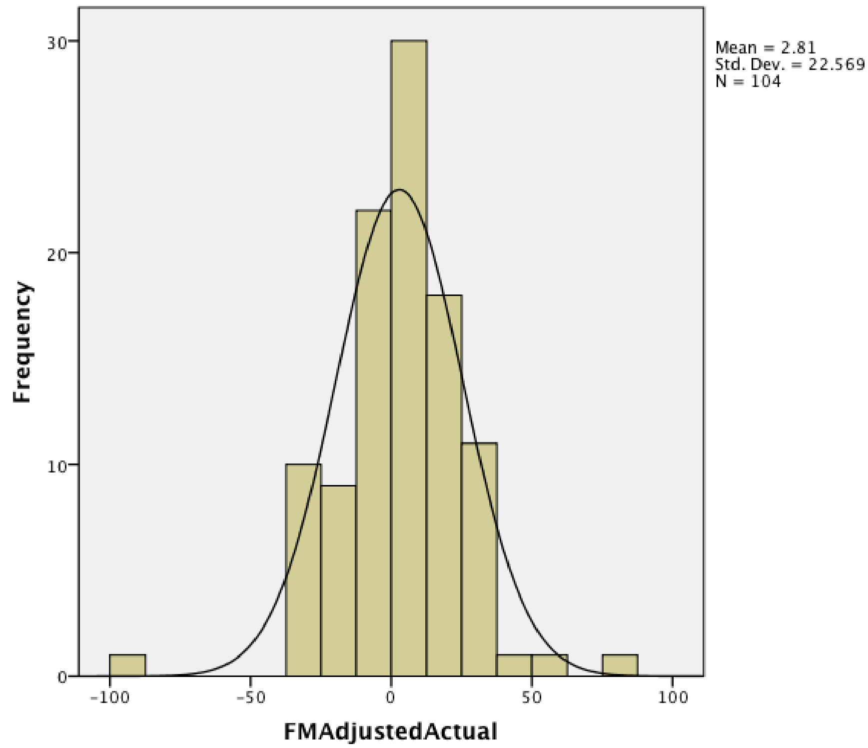


Figure 2. Boxplot: Performance deviation for fine and gross motor subtests



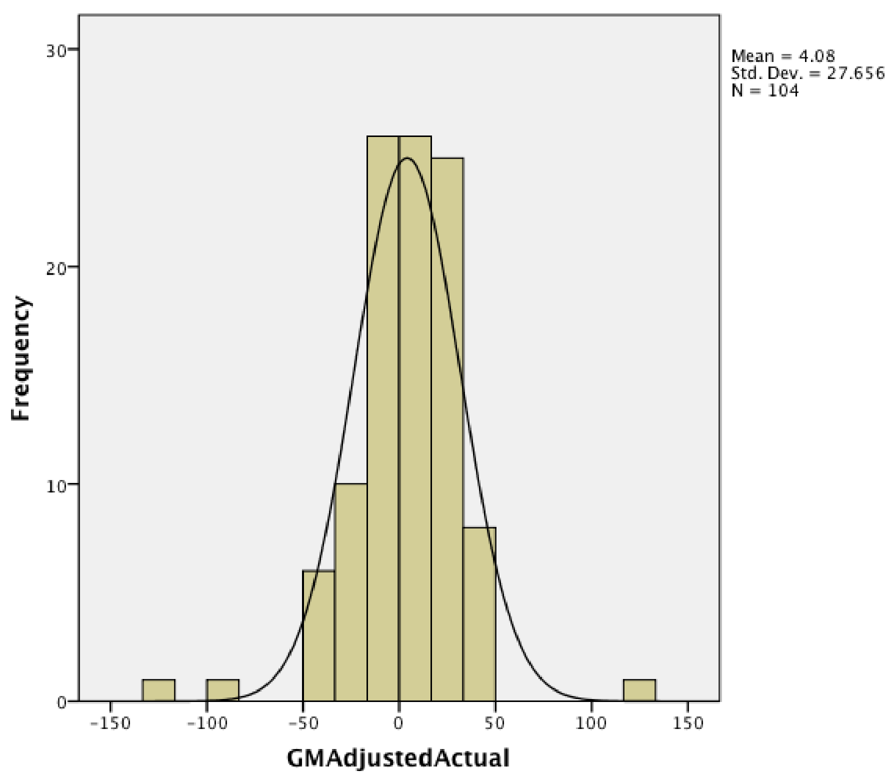


Figure 3. Distribution of age equivalent – adjusted age differences for fine and gross motor

Contingency Table: Gross Motor

	Yes Motor Therapy	No Motor Therapy
Greater than 30 day delay	3	5
Less than or equal to 30 day delay	17	79

Contingency Table: Fine Motor

	Yes Motor Therapy	No Motor Therapy
Greater than 30 day delay	2	5
Less than or equal to 30 day delay	18	79

Note. For Gross Motor, percent agreement = (3 + 79)/104 = 78/8%; percent disagreement = (17 +5)/104 = 21/2%. For Fine Motor, agreement = 77.9%; disagreement = 22.1%. Disagreements were more than three times as likely to be due to the BSID-III indicating no motor delay and the therapist recommending motor therapy (17 or 18 versus 5).

Figure 4. Decision Contingency Tables

Appendix

Data Collection Form : Page 1

ID #	Gender	Length NICU Stay	Dates In NICU	DOB	Birth Gestation	Primary Diagnosis	Birth Wt. (gms)	Date of OT/PT Eval	Bayley Given?	Plagiocephaly	Head Turn Preference

Data Collection Form: Page 2

Chronolgical Age	Adjusted Age	FM Score (Raw)	GM Score (Raw)	Motor Composite	Age Equivalent FM	Age Equivalent GM	% Rank	SD	Follow-up/ Referral	Reason(s)	OT/PT

Acknowledgements

This study was completed under the expert guidance and editorial contributions of Kari Tanta, PhD, OTR/L, FAOTA, and George Tomlin PhD, OTR/L. The author would also like to thank Lauren Adams PT, C/NDT, PCS for her support and guidance and the therapy staff of the setting hospital for their valuable input. Last, the author would also like to thank Professor Matthew Pickard for his contributions to the data analysis portion of this project.