

Efficacy of the START-Play Program for Infants with Neuromotor Disorders

Clinical trial # NCT 02593825

Award # R324A150103

Statistical Analysis Plan

July 17, 2015

i. Sample

Infants with neuromotor disorders will be referred through: early intervention therapists or service coordinators, NICU follow-up clinics, or other sources depending on the specific model of follow up at the various sites. All sites have strong relationships with local sources allowing adequate recruitment (See Resources and Letters of Support Appendix D).

Inclusion criteria:

- Infants enter study between 7-16 months of age, when they are able to prop sit for 3 seconds and maintain their head at least to the level of neutral alignment with their trunk.
- Gross motor delay as reflected in the Bayley III motor subtest >1.0 SD below the mean.
- Neuromotor disorder such as cerebral palsy (CP), or at risk for CP because of extreme prematurity or brain damage that occurred at or around birth, or infants with motor delay of an unspecified origin (no clear diagnosis, but delay as above)
- Eligible infants may meet the U.S. federal criteria for children with disabilities in one of the following categories: *orthopedic impairments, multiple disabilities, other health impairments, or developmental delay.*
- *Minimal movement requirements/Indicators of readiness for change:* Sits with support of arms for 3 seconds after being placed. Exhibits at least some spontaneous movement of arms.

Exclusion

criteria:

- *Medical complications* that severely limit participation in assessments and intervention such as severe visual and congenital/genetic anomalies, uncontrolled seizure disorder.
- Diagnosis other than an unchanging neuromotor disorder (examples: autism, Down syndrome, spinal cord injury, acquired head injury, muscle disorder).
- A child will be excluded if the parents report any of following: 1) if the child has a disability of a progressive nature such as muscular dystrophy; 2) if the child's family plans to move out of the local area within one year from the start of the study; 3) if the child has major surgery planned that might affect physical performance.

Rationale for choosing sites:

- Each site represents a different region of the US: Northwest (UW), northeast (UD), Midwest (Duquesne), and south (VCU).
- The demographics of the regions combined are representative of the US as a whole (average of 4 recruitment sites within 2% of US average for % racial distribution/economic status).
- This project requires experienced primary investigators at each site in order to evaluate, disseminate and expand to other sites around the US in a follow-up project. The team assembled for this project includes the best pediatric early intervention motor and cognitive team that the United States has to offer.
- The assembled team has a proven track record of recruitment/retention for early intervention and family engagement, and strong publication and teaching records for dissemination.

Retention

Attrition is expected to be approximately 10% based on our previous studies; therefore, we plan to recruit 152 infants/families to accommodate dropouts and have 140 complete the study. Strategies to reduce attrition include small honoraria and gifts for participants, intervention and measurement all provided in the home, and flexibility in scheduling to accommodate each family's needs. In addition, thank you notes and gift cards will be provided to the family upon completion of the intervention and each of the follow-up assessments. Our power analyses have considered

this level of attrition and our statistical techniques account for missing data, thus ensuring our ability to complete the complex modeling analyses proposed. If a child is not tracked through the end of the follow-up period or assessment periods are missed but the child remains in the trial, we will utilize full information maximum likelihood (FIML) estimation as a missing data procedure to allow an Intention to Treat (ITT) perspective. Little's (1988) MCAR test and Simonoff's (1988) regression diagnostic procedures will be conducted to determine if missing observations or participant non-response meet the Missing Completely at Random (MCAR) or Missing at Random (MAR) assumptions required by the proposed statistical procedures. Missing data assumed to be at least MAR will be dealt with through FIML without list-wise deletion or imputation methods (Enders, 2001).

ii. Research Hypotheses:

PRIMARY HYPOTHESES

Hypothesis 1: Compared to the BAU group, the START-Play group will show greater improvements from pre- to post- intervention (short-term effect) in:

- a. sitting Variable: Gross Motor Function Measure (GMFM) sitting scores
- b. reaching Variable: Number of toy contacts
- c. problem-solving Variable: IGDI problem-solving score
- d. global development Variable: Bayley III score

Hypothesis 2: Compared to the BAU group, the START-Play group will show greater sustained improvements from pre-intervention to one year since start of intervention (long-term effect) in:

- a. sitting Variable: Gross Motor Function Measure (GMFM) sitting scores
- b. reaching Variable: Number of toy contacts
- c. problem-solving Variable: IGDI problem-solving score
- d. global development Variable: Bayley III score

Hypothesis 3 (Mediation): Improvements in global development will be (at least) partially mediated by improvements in problem-solving as a function of the intervention at the end of the project (long-term proximal effect).

Variables: GMFM scores, toy contacts, IGDI problem-solving and Bayley III scores.

SECONDARY HYPOTHESES

Analysis of moderators

1. *Hypothesis:* Severity of motor impairment will have a significant influence on overall change in IGDI problem-solving, and motor scores, as well as Bayley III scores. Variable: Severity score (see Table 1)

2. *Hypothesis:* Factors external to the child (socialization opportunities [Modified Parent Child Interaction-Dyadic Mini Code scores (Censullo, 1991) (PCI-DMC)], and home environment (HOME Inventory [Bradley, 1993] & SES), will have a moderate effect on change in GMFM, problem-solving, and Bayley III scores.

Further analysis of motor variables

3. *Hypothesis:* The START-Play group will show posture control measures that are closer to normative values than the BAU group from pre to post intervention. This hypothesis tests additional variables quantifying sitting to supplement Primary Hypothesis 1. Variables: Kinematic measures taken from video analysis.

4. *Hypothesis:* Compared to the BAU group, the START-Play Intervention group will show greater improvements in duration of toy contact and object manipulation from pre to post intervention. This hypothesis tests additional variables quantifying reaching to supplement the

Primary Hypothesis 2. Variables: Unilateral and bilateral contact number and duration, and type of contact (i.e. hand open or closed; palmar or dorsal hand surface, manipulation variables).

iii. Research Design

We propose a 3-month randomized trial with repeated measures design plus a nine-month follow-up. The design is longitudinal to best capture changes over time within both the BAU group and the START-Play group, and to distinguish advances in skill due to the specific intervention vs. general development under the usual model. Recruitment of study participants will be continuous throughout the project’s recruitment period, and assignment to condition will occur at the individual child level. It will not be possible to have a control group without intervention because all recruited children will be delayed enough to have mandated early intervention services. The experimental group will receive the START-Play intervention as well as their regular Part C services. Randomization will be stratified by the severity of neuromotor disorder (mild, moderate and severe). This stratification will help to ensure that the intervention groups are balanced. Randomization will be created using a random number generator using blocks of size 4 or 6, and arranged in sealed envelopes by our biostatistician, separately for each site. Numbered group assignment will be dated and faxed to the PI to verify the process.

A scale created in a previous US Dept. of Education NIDRR-funded study of infants with CP helps ensure equal distribution of children with varying severity of neuromotor dysfunction, and has been adapted for this study. The scale is summarized below:

Table 1. Method to Determine Severity of Disorder

Score Received on the Measure	Measures			
	Distribution	Active Movement	MACS	GMFCS
1	1 side	High	1-2	1-2
2	primarily legs	Moderate	3	3
3	all 4 extremities	Low	4	4

Scores from each column are totaled for a final score: 9-12, Severe; 6-8, Moderate; 4-5, Mild.

- Distribution refers to the areas of the body affected by neuromotor dysfunction.
- MACS: Manual Abilities Classification Scale (Hidecker et al, 2009) describes use of the hands and arms
- GMFCS: The Gross Motor Function Classification System (Palisano et al, 1997) classifies a child’s motor involvement on the basis of functional gross motor abilities.

The assessment schedule covers entry month (between 7-16 months of age) and the following 12 months (between 19-28 months of age). Assessments are at Baseline, Intervention (Rx) and Follow-up. Table 2. Assessment Schedule

Assessment number	1	2	3	4	5	6
Timetable (months after study entry)	0	1	2	3	6	12
Sitting Video Reaching Video	Baseline testing prior to intervention	Rx	Mid Rx testing	Final post Rx testing sitting and reaching	Follow-up	Final Follow-up
Standardized Assessment	Bayley III, GMFM, IGDI		GMFM, IGDI	Bayley III, GMFM, IGDI	Bayley III, GMFM, IGDI	Bayley III, GMFM, IGDI

- What: Our primary and secondary measures capture initial impairments and delays in sitting and reaching, and are sensitive to improvements in sitting, reaching and problem solving.
- Where: The natural setting (home) (sitting and reaching video, Bayley, GMFM and IGDI).
- Frequency and Period (Table 2): Measurement visits total 6 over a 12 month period, organized to provide timely assessments without overwhelming families. Repeated measures assess immediate and long term proximal effects on individual and group change over time.
- Who: Blinded, trained assessors who have met criterion standards in assessment procedures (See Appendix C, assessment manual excerpt). The four data collection sites will each send de-identified data videos to a centralized database. Each site has expertise for specific analyses: 1) sitting video data to Duquesne University; 2) reaching data to University of Delaware; 3) IGDI assessment to Virginia Commonwealth University, 4) Reliability for the Bayley and GMFM assessments to the University of Washington, and 5) overall statistical analysis and fidelity management through University of Nebraska Lincoln. The University of Nebraska-Lincoln will create and maintain a secure web-based database interface for all sites to upload finalized data from each site into a central data system, accessible to all collaborators.
- Reliability: Each assessor will be trained and must reach at least 90% agreement criterion with scoring of specifically created videotaped segments of infants for all measures. A second investigator will re-score 20% of the data for each variable to monitor ongoing measurement reliability. For sitting kinematic measures, investigators will be trained to position the infant, monitor behavior etc. to acquire quality data. The sitting kinematic variables are quantified with a mobile video movement measurement application called SIMI Move. A specific, user friendly interface for early infant movement, developed by SIMI Reality Motion Systems, will be available to each site. The sitting measures will be wirelessly sent to the central location, and intra-rater and inter-rater reliability will be assessed across sites to achieve 90% agreement.
- **iv. Power**
- Sample Size: Each of the 4 sites will recruit 38 infants (total $N=152$) to retain 35 infants for a total of $N = 140$ infants in the study. The initial ($N=152$) and retained ($N=140$, $n=70$ per condition) sample sizes were chosen according to feasibility of recruitment in each site as well as considerations of time and resources needed for a 4-year study.
- Power analysis for Primary Hypotheses 1 & 2: We used a simulation-based power analysis (Muthén & Muthén, 2002; Theommes et al, 2010) to estimate the power for the randomized trial with repeated measures design as evaluated with the proposed Linear mixed modeling (LMM; Laird & Ware, 1982; Wallace & Green, 2002) utilizing FIML for attrition-based missing data. *Mplus* was used for simulations and empirical analyses. Simulations were based on detecting significance in the treatment by time interaction during the intervention phase, γ_{11} . The analysis assumed a small intraclass correlation for the slope across infants (e.g., small variance of the treatment effect of 0.05; Raudenbush & Liu, 2000), a two-tailed nominal error rate of $\alpha = .05$, an initial sample size of 152 infants at baseline, and an attrition scenario of 10% loss by the end of the 3-month intervention (based on patterns observed in previous research conducted by the collaborative team). We conducted the power analysis essentially in support of Hypothesis 1 (the short-term effect at the end of the intervention phase) based on the position that Hypothesis 2 will be supported if the initial effect is then at least sustained through the follow-up period which would be evidenced by a non-significant γ_{21} parameter (no differential change). Prior published research by Harbourne et al. (2010) and Heathcock et al. (2008) were used to derive effect size estimates for sitting (GMFM, $d=1.21$) and reaching (toy contacts, $d=1.32$), respectively. Pilot data from Galloway & Lobo and Harbourne were used to derive

effect size estimates for global development (Bayley III, $d=0.43$) and problem solving (IGDI, $d=1.11$), respectively.

- Based on our previously stated assumptions and the cited pilot data and prior research, power analyses suggest that the targeted sample size will be sufficient to detect the reported effects with greater than 80% power for three of the four primary outcomes. With $N=152$ initial participants, power to detect mean group differences in slopes during the intervention period is >99% for sitting, 93% for reaching, and >99% for problem solving. As these three outcomes are estimated to be over-powered based on pilot/cited data, given the projected initial sample size of $N=152$ we would be able to detect effect sizes as small as $d=0.48$ for sitting, $d=0.66$ for reaching, and $d=0.56$ for problem solving. Based on the effect size derived from pilot data, power to detect mean group differences in global development improvement during the intervention period at the $d=0.43$ level is only 48%; however, given the projected sample size, we would be able to detect an effect size of $d=0.64$ with 80% power.
- *Power analysis for Primary Hypotheses 3 & 4:* Continuing to use the variability and mean differences from the cited prior research and pilot data as estimates of paths A and D in Figure 2, we conducted an empirical power analysis based on the SEM model presented in Figure 2 using an initial sample size of $N = 152$ infants, and assuming a moderate correlation between the measures at baseline ($r = .32$, explained variance = 10%). With these assumptions, we will have 80% power to be able to support the hypotheses of partial mediation of intervention-driven change over time in global development by intervention-driven changes in sitting and reaching and problem solving if the standardized parameter estimate of the mediator(s) on change in global development (paths B (Hyp. 3) or $C_1 * C_2$ (Hyp. 4) in Figure 2) is at least $\beta=.237$, which corresponds to a medium effect.
- A complementary power analysis perspective for evaluating Hypotheses 3 and 4 follows MacCallum et al.'s (1996) guidelines for determining adequacy to test the fit of a model, which indicates whether the necessary parameters/effects are being considered. Evaluation of model fit is a prerequisite for evaluating parameters within the model itself. Accordingly, a hypothesis of *close model fit* can be tested by assuming the Type I error rate to be 0.05, a $RMSEA \leq 0.05$ as indicative of *close fit*, and a $RMSEA > 0.10$ as indicative of poor fit. Given the proposed model in Hypothesis 3, $df_{min}=51$, $N=152$, $RMSEA_{null}=0.05$, $RMSEA_{alt}=0.10$, and $\alpha=0.05$, the power to detect close model fit for Hypothesis 3 is >.99. Given the proposed model in Hypothesis 4, $df_{min}=115$, $N=152$, $RMSEA_{null}=0.05$, $RMSEA_{alt}=0.10$, and $\alpha=0.05$, the power to detect close model fit for Hypothesis 4 is .96.
- In summary, the initial sample size of $N=152$ and retained sample size of $N=140$ (70 per group) is appropriate for the study and will likely yield power to test each primary hypothesis at a medium effect size or larger at the $p < .05$ level, with a power of at least .80. The structure of this proposal ensures that we can realistically produce the first randomized efficacy trial of sitting and reaching with young infants within the funding period.

v. Measures [Outcomes]

Rationale for using multiple measures of developmental skills

The use of standardized tests in isolation is not optimal for infants with neuromotor disorders because of the large increments of change in observable motor skills needed to change the score on any one item, and the complexity of child behavior. Furthermore, our research has shown that standardized assessments like the Bayley III focusing on the discrete performance of tasks are not consistent for identifying and classifying delays in the first two years of life and are poorer predictors of future cognition relative to assessments that involve perceptual-motor

exploration and problem-solving (Lobo, Paul, et al, 2014; Lobo & Galloway, 2013). However, we want a means to relate our primary measures to other developmental studies. Therefore, we are using a blend of standard tests (Bayley III Scales of Infant Development, individual change assessments (Individual Growth and Development Indicators [IGDI]); Gross Motor Function Measure [GMFM]) and video coding measures to quantify our secondary variables for sitting skills and reaching.

Mindful that specific outcomes are considered critical to children becoming active and successful participants across a variety of settings according to the US Department of Education (OSEP), we acknowledge these outcomes as related to our measures: 1) Positive social-emotional skills and social relationships; 2) Appropriate behaviors to meet needs.

The proposed START-Play intervention relates directly to these critical and functional outcomes. The parent-child interaction measure provides information on the child's progress on outcome 1. The problem-solving IGDI and the GMFM provide information on the child's progress on outcome 2. We are also using the secondary measurable outcomes below, which we have used in previous funded studies, to detect small increments of change and quantify progress of skill in sitting and reaching.

Because this is a longitudinal study, the infant's skills in reaching are expected to change considerably. Although we have based our power analysis on toy contacts, we want to completely describe the advances in reaching as the children gain skill. Thus object manipulation, type of hand contact, and other measures will be coded advanced skill.

Specific Measures

1) Primary Measures for Sitting and Reaching

These variables quantify the initial functional delays in these skills and changes during the Intervention and Follow-up periods.

- GMFM sitting section

The Gross Motor Function Measure (GMFM, Russell et al 1993) is the international standard outcome assessment tool for quantifying change in the gross motor abilities of children with neuromotor disorders (inter-rater reliability, $r=.97$, Russell et al 2000). Each sections raw score and % of the section completed increase as abilities improve.

- Number of toy contacts

Data Collection: Infants are provided six 30 second trials to reach for an object sitting upright on the floor under 3 different levels of support for a total of 15 trials One of three levels of support are used: at the waist, at the hips, or no support. Trials begin when infants are visually attending to the object. Infants are presented with a graspable toy in midline at chest height 75% of arm length. Two synchronized video cameras assure a clear view for behavioral coding.

Data Analysis: Coders will determine number of toy contacts (and secondary variables, see section below) by coding the videos. Contact occurs when any surface of the hand contacts the object. Coders are blind to group assignment and study purpose. Coding reliability is assessed on 20% of the sessions and is based on a comparison of agreements and disagreements from each visit. Based on previous work, coders will maintain inter- and intra-rater reliability >90%.

2) Primary Measure for Problem-Solving

- Individual Growth and Development Indicators (IGDIs)

The Infant and Toddler IGDIs are brief psychometrically sound measurements of young children's abilities (Carta et al, 2002) across 4 outcome areas. We have found the Early Problem Solving Indicator (Greenwood et al, 2006) (EPSI) to be sensitive and suitable for assessing infants when assessors are well-trained (>90% agreement). It defines problem solving as consisting of

visual exploration, object manipulation and memory. Previous studies show good reliability ($r=.93$) and validity, and its usefulness in documenting change over time. To test, the child is videoed for 6 minutes when presented with standardized toys. Key skill elements are coded for frequency from video. Frequencies are summed, weighted as to complexity of the behavior, and divided by time for a total rate. We will use the certification process online to insure that IGDI testers and coders attain a satisfactory level (90%) of reliability. Members of this consortium have done pilot work to test the problem solving IGDI on children with neuromotor disorders, and have created standardized adaptations for children with limited movement.

3) Primary Measure for Global Development

- Bayley Scales of Infant and Toddler Development—Third Edition (Bayley III, Bayley, 2006), The Bayley III is a well known norm-referenced test designed to assess five key developmental domains: cognition, language, social-emotional, motor and adaptive behavior. Videotapes of this test will be scored to reach reliability of 90% agreement at the UW site.

4) Secondary Measures: These variables quantify factors related to the primary measures above. Secondary Measures for Sitting related to impairments in postural control:

Trunk alignment as an infant develops sitting posture has long been used as a clinical measure of postural control (McGraw, 1943; Bly, 1983; Harbourne et al, 1993). Movement and posture in sitting will be sampled monthly and quantified with kinematic measures from video.

Instrumentation: Quantification of sitting posture data will occur via a video taken from the side view of the infant (Harbourne et al, 1993). Data will provide:

- angle of forward bending during independent sitting,
- angular velocity from vertical to self-supported resting position, and
- angular displacement from vertical supported position to self-supported sitting position.

These variables are indicators of the control the infant is developing for maintaining upright sitting. Standardized positioning of child and camera will allow accurate use of the SIMI Move application to quantify the trunk control in sitting (90% agreement)(Surkar et al, 2013).

Data Collection and Analysis: The infant will be placed in a sitting position on the floor, and supported by the examiner; the examiner then removes their hands, which have been supporting the trunk. Measurement is taken from the point of the initial supported position to the end position independently assumed by the infant, and is quantified by degrees. Data are collected for 3 trials while the child attempts to maintain sitting. *See Figure 8 Appendix B.*

Secondary Measures for Reaching (documents impairments in reaching and object exploration):

These variables are coded from video to further quantified reaching:

- Age of reach onset and number infants that are reaching at each age. An infant is ‘reaching’ the first session he contacts the toy more than 5 times in 3 min of trial time (Lobo et al 2004).
- Percent of time each hand is in contact with the toy (aka Contact Duration).
- Number of toy contacts with both hands
- Hand position at initial contact and during contact: open/closed hand, ventral/dorsal surface.

Secondary Measures: Child and Family Characteristics

Child characteristics (Severity, Cognitive, Age, Health) will be recorded, respectively, from previously described Severity Index, Bayley III scores, and a Demographic questionnaire and Health questionnaire, which are both completed by parents. Environmental factors include: 1) Parent/child interaction (PCI-DMC) See Figure 9, Appendix B., 2) Home environment (HOME scale [Elardo & Bradley, 1981]), 3) SES, 4) program attendance/participation (visit counts), and 4) overall amount of services (Therapy Services questionnaire). The Home Observation for Measurement of the Environment (HOME) Inventory – Infant/Toddler version assesses the home

environment including adult/child contact, social/cultural experiences, physical environment, and access to toys. SES will be measured on a Demographic questionnaire completed by parents. Program attendance will be recorded by the number of accomplished visits, documented by the therapist visiting the home. Amount of therapy will be recorded via the Therapy Service questionnaire completed by parents (see Appendix B).

vi. Fidelity of implementing the START-Play intervention

Our fidelity approach will determine the degree to which all elements of the START-Play intervention are delivered across treatment and control conditions (Knoche et al, 2010). We will assess dosage, adherence, quality, participant responsiveness, and program differentiation (Berkel et al., 2011; Cordray & Pion, 2006; Dane & Schneider, 1998; O'Donnell, 2008) with three primary measures: (1) START-Play Home Visit Coding Guide (adapted from the Home Visit Observation Form; McBride & Peterson, 1997) (see Appendix B); (2) Therapy Services Form (adapted from Westcott & Effgen, 2011; Appendix B); and (3) Home Activity Log (Appendix B). Both the experimental and BAU groups will be measured using these instruments. Relevant fidelity metrics for dosage, adherence and quality (or some combination) will be derived in consultation with our team from the Statistics and Research Methodology Unit of CYFS, allowing us to account for fidelity as a covariate in the analysis of treatment effects.

Two components of fidelity are particularly relevant in our investigation. Both what the therapist does during the home visit sessions, and the intervention as carried out by the family will be measured. **Adherence** to specific features of the START-Play intervention (type of practice) and **quality** of implementation will be measured in both treatment and control groups with the START-Play Home Visit Coding Guide (Appendix B). This guide contains both adherence items (1-10) and quality items (items 11-16) to yield an adherence score and a quality of implementation score. This data will be collected via videotape by an assistant trained for data collection, and then coded in the lab by a trained and reliable observer. One session per month during the three intervention months will be observed. This measure yields a total adherence score indicating the percentage of criterion steps (percent of items 1-10 addressed during session, yes/no) implemented by the therapist, and a total quality score reflecting accuracy of implementation (average of times items 11-16 are performed to reflect a rate of implementation).

Dosage of both amount of services received (i.e., START/EI and outpatient) as well as specific tasks (amount of practice) will be recorded for both groups. An attendance log recorded by the interventionist will document number of home visits completed by both groups. This data is collected in EI programs, and should be available in the BAU group. Additional services for the infants outside of the START-Play intervention and their EI intervention (e.g., outpatient therapy) will be counted on the Therapy Services Form to quantify dosage of motor intervention.

Implementation of the intervention by families (**participant responsiveness**) will be documented at each assessment time for both groups using the Home Activity Log (see Appendix B), measuring how frequently families or caregivers implement strategies learned during the home visit sessions when therapists are not in the home. Families will briefly document how often and what type of activity suggestions are carried out in the home

For the START-Play group, fidelity of implementation will also be maintained by using the START-Play Intervention Checklist that identifies essential criteria of the intervention (see Appendix C). At the end of each session, the interventionist will check off each objective addressed. The interventionist will record notes, and parent contributions to each session. We will video three sessions per therapist at the start of the study and the site PI will observe and code using the intervention checklist. 90% agreement between the site PI and therapist self-check is

targeted. If 90% agreement is not attained, the site PI will re-train the therapist and repeat the assessment. Upon attaining 90% match with PI in performance, one video per month will be checked by the site PI to maintain intervention fidelity. The project PI and each site PI will be available for consultation to each therapist as needed. Monthly meetings will be held via on-line video conference to assure fidelity by reviewing strategies, activities, and problems.

vii. Comparison (BAU) group

The comparison group will receive their usual early intervention without any interference. They will receive the same assessments following the same schedule as the infants in the intervention group. Although the children in the comparison group may be working on some of the same goals as the intervention group because they are at a similar point developmentally, no training or support related to the START-Play program will be provided.

Early intervention programs characteristically provide variable services to each child. Therefore, it may be that the control group receives services at least partially similar to the START-Play group. The START-Play Home Visit Coding Guide will be utilized monthly during the three-month intervention period to document and describe differences or similarities in intervention programs (**program differentiation**) between the BAU and START-Play groups. In addition, the Therapy Services Form will describe the frequency and amount of intervention received by each child, and will be used to quantify and evaluate additional services. The overall composition of both groups is expected to be equal because of the randomization process.

Anticipated differences between the START-Play intervention and the BAU intervention are summarized in the following Table 4.

Table 4. Comparison of START-Play to Business as Usual Early Intervention program

	Business-as-usual*	START-Play	Difference
Frequency	Once weekly to once monthly	Twice weekly	Increased frequency of interaction
Intervention Model	Transdisciplinary or interdisciplinary- may be teacher, PT, OT or Speech	Short term focus with expert in motor (physical therapist) in partnership with parent for activities below (<i>Appendix C</i>)	Increased focus for caregiver on understanding motor and linkage to problem-solving
Content of Session	Directed to Parent: (25%) <ul style="list-style-type: none"> • Parent Training 25% Directed to Child: (55%) <ul style="list-style-type: none"> • Task Guidance w/ PT 25% • Postural Control & Strengthen w/ PT 20% • Tolerate Positions 10% • Equipment/Stretch w/ PT 10% 	Directed to Parent: (35%) <ul style="list-style-type: none"> • Analysis with parent & setup 20% • Parent-led activity 15% Directed to Child: (40%) <ul style="list-style-type: none"> • Child-led activity with sitting /reaching 25% • Expand child-led activity - variability 15% Readiness to learn related: (25%) <ul style="list-style-type: none"> • Link motor acts to problem-solving and cognition 25% 	Parent/caregiver taught all aspects, not just tasks, but also the “why” behind the action, connections to previous and later skills, problem-solving links, cognitive outcome expected for each action

	Readiness to learn related: (10%) • Cognitive - < 10%		
Example Content	<ol style="list-style-type: none"> 1. Practice sitting balance reactions on a ball (isolated motor task) 2. Therapist provides supportive seat to constrain trunk, and suggests presenting toys on tray in front of infant and modeling use of toy 	<ol style="list-style-type: none"> 1. Select activity for motor based problem solving – finding hidden toy. Child encouraged to shift weight, re-orient to look behind/under/in containers, thus building sitting balance in the service of spatial understanding 2. Dynamic low support sitting allows child to re-orient and gain spatial understanding; multiple options for variable sitting support depending on problem-solving task. 	<ol style="list-style-type: none"> 1. Cognitive construct is selected first and is primary; movements are built around the cognitive construct 2. Parents taught that chair is passive and not variable. Multiple options, with minimal support needed, allow exploration and link motor to problem solving

*Percentages indicated were obtained from an ongoing consortium study of standard of practice by early interventionists at the 4 planned sites

Description of intervention

In addition to the differences noted above, and the focus on early skills of sitting and reaching, the approach of the intervention is notably different from that of a business-as-usual early intervention visit. A BAU visit might consist of the therapist trying and then showing the parent routine motor activities. These handling routines often consist of holding the infant so that trunk support is reduced gradually in routine activities such as holding the child, dressing, carrying and feeding, and are end-points without linkage to problem-solving or cognitive tasks. The START-Play group will utilize a perceptual-motor approach (Tscharnuter, 2002), which uses self-initiated goal-directed movements to build and bolster orienting and attending to objects, while understanding basic relationships of cause and effect through manipulation and focused attention. The specific intervention depends on the skill of each child. Generally, activities focus on helping the child learn to attend to significant environmental information, such as pressure against surfaces, which can be correlated to forces useful for controlling posture and movement. Unlike passive positioning or movement in BAU approaches, our approach encourages active, child-initiated movement to achieve environmental goals (Ziegler et al, 2010), and learning to solve problems linked by movement and manipulation of objects, which then scaffold cognitive skill; the focus is not on a “normal” movement pattern. Parents learn to discover and problem-solve motor challenges, and link small motor changes to cognitive skills, increasing opportunities for the child to link motor and cognitive areas. See Appendix C.

viii. Mediating and moderating variables

Skilled sitting and reaching are expected to affect problem-solving skill, which is expected to influence global development. Problem solving skill is then a mediator of the effect of change

over time between motor skill and cognitive skill. Testing for mediation will take place by 1.) showing that global development change is influenced by change in sitting and reaching and 2.) change in reaching and sitting are linked to problem solving, 3.) change in problem solving is linked to change in global development and 4.) the effect of change in sitting and reaching on global development is at least partially explained by change in problem solving.

Other factors which may be moderators of the final outcomes are 1) severity of neuromotor disorder, 2) health status, 3) age at entry, 4) cognition at entry. Factors external to the child which may moderate outcomes are: 1) Socioeconomic status, 2) compliance or attendance to the program visits, 3) the home environment (HOME), 4) overall amount of service provided to the child, and 5) fidelity of intervention implementation. We expect that these moderating variables will not be different between groups because of our randomization process, however tests for initial differences between groups will verify this. In addition, testing for moderation will take place after the primary research questions are addressed. Variables will be tested individually first and later in combination to account for increases in explained variability.

ix. Data Analysis: Evaluation of Specific Aims and Hypotheses

In this randomized trial, we propose a repeated measures design where individual children serve as the unit of randomization. This experimental study will involve up to 38 children from each of 4 geographic sites resulting in up to 152 total participants. Participating children will be assessed on up to 6 unevenly spaced occasions per the assessment schedule in Table 2 with the metric of time being the number of months since the beginning of the study. Linear mixed modeling (LMM; Laird & Ware, 1982; Wallace & Green, 2002) is an appropriate approach in this study due to presumed individual child differences in rate of change and correlated outcomes due to hierarchical nesting, the uneven spacing of measurement occasions, and the potential for missing data due to attrition. The same fundamental model will be used in simultaneously testing Hypotheses 1 and 2, with models differing only in the particular outcome included in the model. Hypotheses 3 and 4 will require parallel process growth modeling (see Cheong et al., 2003), which is a synthesis of LMM and structural equation modeling (SEM). The discussion of analytic methods for Hypotheses 1 and 2 will be presented in a LMM format, and the discussion for Hypotheses 3 and 4 will be in SEM format. SAS PROC GLIMMIX (currently v13.1) will be used for LMM analyses, and Mplus (currently v7.2) used for SEM analyses.

Prior to evaluation of specific study aims and hypotheses, all variables at all measurement occasions will be described by numeric summaries and visual displays (such as histograms and boxplots) to assess for normality and test other prerequisite modeling assumptions.

Evaluation of Primary Hypotheses 1& 2. A piecewise LMM with repeated measures will be used to simultaneously address the primary research hypotheses: (1) compared to the BAU group, the START-Play group will show greater improvements in sitting, reaching, problem-solving and global development from pre- to post-intervention (short term effect); and (2) compared to the Business as Usual (BAU) group, the START-Play group will show greater sustained improvements from pre-intervention through one year since the start of intervention (long-term effect). The effects of interest will be the child-level *time by group* (START-Play vs. BAU) interaction evaluated at (a) the end of the third month (Hypothesis 1 - after completion of the intervention period), and (b) the long-term follow-up phase through kindergarten (Hypothesis 2). Tested simultaneously, these interactions will indicate whether the change in outcomes due to participation in the START-Play intervention condition differs significantly from any change in outcomes that occur with BAU. Primary outcome measures (GMFM) sitting scores, toy contacts, IGDI (problem solving), Bayley III) will be analyzed separately so all measures can be clearly

assessed for the full potential of information. Appropriate error-rate control is discussed in a later section. Severity of neuromotor disorder (stratification variable) and fidelity of implementation will be added as covariates, and child-level characteristics, such as gender and ethnicity, also may be included in the analysis as child-level covariates. Due to the small sample size ($J = 4$) and the convenience sampling of sites involved in the study, site will be included as a *fixed* classroom level effect rather than an additional hierarchical level or random effect. Our analytic model is presented as a piecewise LMM where Y_{it} is the outcome variable score at time t for child i , TRT_{0i} is a contrast variable denoting the intervention status for child i (i.e., 0=BAU, 1=START-Play), RX_{it} is the passage of time during the intervention phase of the study from baseline to the end of the third month, and LT_{it} is the passage of time from the end of intervention through one calendar year since entering the study. While presented here as linear effects within each study phase (RX and LT), we will also explore any potential curvilinearity as well. The second line of the equation reflects the remaining fixed covariate effects ($CTRL_{0i}$ =neuromotor disorder severity, fidelity;

COV_{0i} =child covariate [i.e., gender]; and S_{0i} =site). *The model can be expanded to consider additional covariates*

$$Y_{it} = \gamma_{00} + \gamma_{01}TRT_{0i} + \gamma_{10}RX_{it} + \gamma_{20}LT_{it} + \gamma_{11}TRT_{0i}RX_{it} + \gamma_{21}TRT_{0i}LT_{it} + \gamma_{02}CTRL_{0i} + \gamma_{03}COV_{0i} + \gamma_{04}S_{0i} + u_{0i} + u_{1i}RX_{it} + u_{2i}LT_{it} + \varepsilon_{it}$$

such as Bayley score at baseline by adding additional parameters to this portion of the model. A random intercept (u_{0i}) and random slopes ($u_{1i}RX_{it}$ and $u_{2i}LT_{it}$) will be included to capture individual differences in outcome levels at entry into the study, and change through each phase of the study (intervention and follow-up). Finally, we will include a residual term (ε_{it}). We expect that γ_{01} , the mean difference between conditions at baseline, will not differ significantly from zero, reflecting no initial group differences and an effective randomization. This parameter also serves to statistically control for any differences at baseline (significant or otherwise) that may be observed despite the random assignment to condition. The parameter, γ_{11} , as the difference between condition slopes through the end of the intervention, is expected to be significantly greater than zero, reflecting a larger rate of change in outcomes as a result of START-Play participation; and γ_{21} as the difference between condition slopes during follow-up will either be not significantly different from zero, reflecting maintenance of the net intervention effect, or significantly greater than zero, reflecting a continued improvement in outcomes due to the START-Play intervention. Estimates of effect sizes with confidence intervals will be provided.

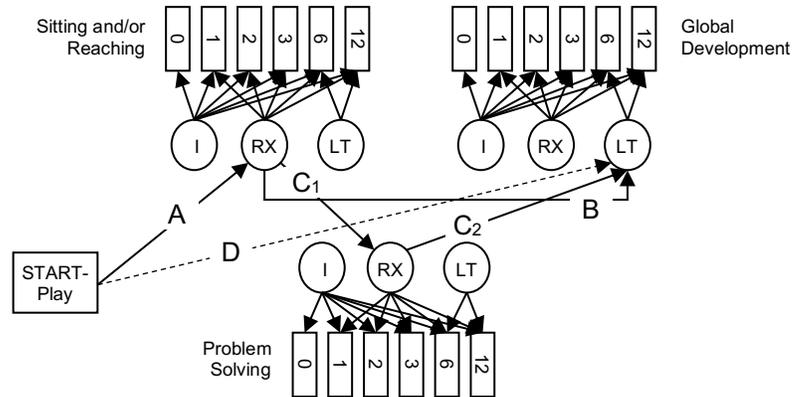
Control of the family-wise error rate (FWER) is necessitated by the presence of multiple outcome measures as discussed in the prior **rationale for using multiple measures of developmental skills**. We define the family of tests as containing the four primary outcomes as described in the Measures section. We will control FWER by conducting the permutation step-down procedure (Westfall & Young, 1993) within a repeated measures design (Lix & Sajobi, 2010) allowing for estimation of intervention effects while controlling for correlation between outcomes due to the common theory of change construct.

Evaluation of Primary Hypotheses 3 & 4:

A parallel process growth modeling approach as implemented in the SEM framework will be used to simultaneously address the remaining primary research hypotheses: (3) improvements in sitting and reaching as a function of the intervention will be associated with increases in global development at the end of the project (long term effect); and (4) improvements in global development will be (at least) partially mediated by improvements in problem solving as a function of the intervention at the end of the project (long-term effect). Because the global development and problem solving constructs occur over time and are characterized as improvements (i.e. changes), we consider the mediational model as a SEM (see simplified model in Figure 1) that allows multiple simultaneous direct and indirect paths while modeling growth (Cheong, 2003). This approach will allow us to assess possible mediating effects of improvements in sitting, reaching, and problem solving through the intervention period (RX) on the relationship between intervention participation (START-Play) and improvements in global development through the one-year follow-up (LT). This model will be evaluated for the significance and directionality of included path coefficients and overall model fit. It is expected that the indirect effect of the intervention (derived by tracing solid paths $A*B$, and $A*C_1*C_2$) on child outcomes will be statistically significant and the direct effects of the intervention (dashed path D) will also remain significant, thus demonstrating partial mediation. Hypothesis 3 will be evaluated by comparing the magnitude of the $A*B$ indirect effect to the direct effect, D. Hypothesis 4 will be evaluated by comparing the magnitude of the $A*C_1*C_2$ indirect effect to the direct effect, D. Partial mediation is an appropriate hypothesis as the intervention is expected to have a substantial effect on problem solving, but there are other factors beyond problem solving, sitting and reaching that contribute to improvements in global development that are represented by the direct effect. By hypothesizing that assignment to condition (START-Play) leads to proximal changes in the mediator(s) (RX) which then lead to distal effects in the outcome (LT), we also meet the requirement of temporal precedence necessary for strong mediational inferences.

Evaluation of Secondary Hypotheses 1 & 2: Improvements in sitting, reaching, problem-solving and global development will be moderated by (1) internal variables: the severity of neuromotor disorder, health status, entry age, cognitive status at entry; and (2) external variables:

Figure 2. Mediated Effects Model for Hyp. 3 & 4



Note: RX = intervention phase; LT = follow-up phase; START-Play = assignment to condition; A,C = direct effects of the intervention on proximal & distal outcomes, B₁ & B₂ = effect of mediator(s) on other outcomes

socioeconomic status, compliance/attendance to home visits, the home environment (HOME), and overall amount of service to the family and child. Evaluation of these variables as moderators of the effect of the intervention on change in primary outcomes (sitting, reaching, problem solving and global development) will be done by modifying the LMM model equation used in evaluation primary hypothesis 2. Two three-way interaction terms, $TRT_{0i}RX_{it}MOD_{1i}$ and $TRT_{0i}LT_{it}MOD_{1i}$, where MOD_{1i} represents an individual moderator, will be added to the model to capture the intervention by time by moderator interaction. The simple effect of the moderator, MOD_{1i} , will be included as well as relevant 2-way interactions. All continuous variables will be mean-centered prior to forming cross-product interaction terms, and all categorical variables will be dummy coded prior to forming interaction terms. Each potential moderator will first be entered individually, and later in combination, with each primary outcome tested separately.

Evaluation of Secondary Hypothesis 3: Relative to infants in the BAU group, infants in the intervention group will show greater improvements in sitting control both post-intervention and at long-term follow-up. Evaluation of this secondary hypothesis will utilize the same LMM framework presented for evaluating primary hypotheses 1 and 2, with the exception that the outcome measures will be trunk forward bending, angular displacement, and velocity of trunk.

Evaluation of Secondary Hypothesis 4: Infants in the intervention group will show greater improvements in reaching from pre- to post-intervention than the infants in the BAU group. Evaluation of this secondary hypothesis will also utilize the same LMM framework presented for evaluating primary hypotheses 1 and 2, with the exceptions that (a) the outcome measures will be unilateral and bilateral contact number and duration, and type of contact (i.e. hand open or closed; palmar or dorsal hand surface), and (b) that only data through the end of intervention will be used (i.e. there will be no LT_{it} term in the model). As these outcomes are categorical, we will utilize the capacity for SAS PROC GLIMMIX to implement a *generalized* LMM framework by including a logit link function and a cumulative or generalized logit link function to model dichotomous outcomes and polytomous outcomes (more than 2 categories), respectively.