Official title: Integrated Oxytocin and Nonverbal, Emotion Recognition, and Theory of Mind Training for Children With Autism Spectrum Disorder

NCT number: NCT02918864

Document date: 2/8/2021

Statistical Analysis Plan

Data Reduction and Outcome Variables. Safety and tolerability will be evaluated by comparing group differences on total counts and domains of the AE form and ratings on parent satisfaction surveys respectively. Efficacy analysis will follow analytic procedures used to derive social behavior impairment (SBI) and social cognition (SC) composites for the Seaver-NETT RCT16. First, principal components analysis (PCA) with promax rotation will be conducted on the primary social behavior impairment measures, the Nonverbal Communication and Social Relations subscales of the Children's Communication Checklist-2 (CCC-2) and the total score of the Griffith Empathy Scale (GEM). A separate PCA will be conducted on the primary social cognition outcomes, the Reading the Mind in the Eyes Test (RMET) and Diagnostic Analysis of Nonverbal Accuracy-2 (DANVA-2). The PCAs are used to confirm that these measures contribute to previously identified constructs in a field with no viable outcomes for higher-order core social impairments as well as to minimize floor and ceiling artifacts, effects of variability in response, and other sources of measurement error.

Prior to the PCAs, missing, invalid, and/or incomplete data will be removed. For SBI, observations missing one or more items on either CCC-2 subscale and/or greater than 30% of GEM items will be removed. For SC, observations with less than 50% of the responses (due to administrative error) on the RMET and/or DANVA2 will be removed. Provided these measures have high and near comparable component loadings on a single factor for each domain, thus confirming previous findings, they will be combined for separate composite scores for social behavior impairment (SBI) and social cognition (SC). To create the SBI composite score, total score on the GEM is reversed so that the higher scores reflect greater impairment as on the CCC-2. To create the SC composite score, percent correct was calculated for the RMET and DANVA-2 since they differ in total number of items and to adjust the denominator when all items were not presented due to administrative error. Subsequently, for each measure, scores will be standardized using the sample mean (and standard deviation) at baseline to account for change from baseline analyses, different and wide ranges of scores, and to ensure that composite scores across domains can be easily compared on the same scale (i.e., z-score). Following outlier detection (see below), standardized scores for each domain (SBI, SC) will then be averaged together to generate final composite scores at each time point. As such, a change in composite score reflects a change from baseline.

Outlier Detection. Individual scores from measures that are either influential points or outlier values (i.e., based on Cook's distance > .5 and/or Bonferroni-corrected outlier based on r-student residuals, respectively) and plausibly invalid (e.g., due to clinical reasons or administrative errors) will be removed. Composite scores will then be re-standardized and remaining extreme outliers (>3 SD above the mean for a specific timepoint) will be removed if they change the model significantly (>10% change in estimated values).

Data Analysis. *Primary:* Composite scores of both primary outcomes (i.e., SBI and SC) will be used as the response (i.e., outcome) variable in a series of mixed-effects linear regression models using the *lmer* package in *R Studio* to assess the effect of treatment condition (ION-ASD vs. Facilitated Play), time (e.g., change from baseline after 12 weeks), and condition X time interaction. Across all models, the individual participant will be included as the random effect. Age, verbal IQ, ADHD and generalized anxiety (from the Child & Adolescent Symptom Inventory-5 [CASI-5] T-scores) will be entered into the model to be examined as potential covariates and/or moderators to outcomes as in prior work. We will estimate effect sizes (standardized Betas) and confidence intervals (95%) acknowledging the likelihood that analyses may not permit definitive conclusions given the likely size of the confidence intervals. Finally, if the statistical assumptions of the model are not met (e.g., homoscedasticity, normality, linear relationships), generalized linear models with appropriate link functions or potential transformations of predictor variables will be assessed.

Secondary Outcomes. The linear mixed-effects models will also be used on secondary outcome measures: SRS T-score, mean Caregiver Strain Global rating scale. Responder (i.e., Clinical Global Inventory-Improvement [CGI-I]) analysis will be conducted using either a Chi-squared or Fisher's exact test (when observations are small such that >20% of cells have *expected* count <5) on the number of responders (i.e., CGI-I scores of 1-2) versus non-responders (i.e., CGI-I scores of 5-7) by group.