

Cover Page

Statistical Analysis Plan

Study Title: Longitudinal Follow-up of Brief Parenting Interventions to Reduce Risk of Child Physical Maltreatment

Identifier: NCT04059185

Document date: 04/18/2024

Statistical Analysis Plan

The power analysis was based on the continuous primary outcome scores at 3-month follow-up. Assuming a significance level of 0.05 in two-tailed tests for the interaction between treatment group and visit type (baseline and follow-up), we would have at least 90% power to detect a standardized effect size of 0.33, if we enrolled 300 participants per group. Conservatively we allowed for a modest amount of autocorrelation of 0.30.

The reported results will follow an Intent to Treat analysis. For the baseline to 3 month follow up analysis, we will fit linear mixed effect models for each of the scale outcomes: Attitudes towards Spanking Scale (ATS), Positive Parenting Practices scale (PPP) and Child Emotional Maladjustment (CEM). The predictors in the models will be group membership (TP2, PN, and Control), visit type (baseline and follow-up) and their interaction along with random intercepts for subjects. We will also reported contrasts among treatment groups and visit type based on post-hoc least squares means analyses, using a Benjamini and Hochberg correction to adjust for false discovery rate for multiple comparisons. As a robustness check, we will use difference scores as the outcomes in linear regression model. We will use Full Information Maximum Likelihood to address missingness. Generalized Estimating Equations with a multinomial outcome will be used for the categorical outcome. We will use multiple imputation to maximize use of all available data.

For the 4 year follow up (4Y) analysis, we will fit linear mixed effect models for each of the scale outcomes: ATS, PPP (total and 4 subscales) and CEM (total and 2 subscales). The predictors in the models will be group membership (TP2, PN, and Control), visit type (baseline, 3M, 4Y), and their interaction. Models will also include random intercepts for participants. Generalized estimating equations (GEE's) will be used for the nominal multinomial outcome of spanking behavior, accounting for within-person correlation across timepoints.

Multiple imputation using chained equations will be used to address missingness and unusually delayed 4Y visits due to the COVID-19 pandemic (number of imputations = 100). Specifically, a new column will be created for "4-year-follow-up" responses on all primary outcomes and filled in with the participant's true response if they were visited within 5.1 years of their baseline visit (i.e. if their follow-up time was below the third quartile for the sample follow-up time); Otherwise, this value will be set as "missing" and therefore imputed using the multiple imputation procedure. The imputation model will include participants' actual long-term response values and their observed time lag to long-term visit from baseline interview date, in addition to the primary measures used in the study at all time points, demographic information, and intervention group.

For the RSA outcome measures, linear mixed effects models will be employed to model the effect of the intervention condition on task specific RSA withdrawals. For these questions, children and parent data will be analyzed separately, and the specified model will be identical for both populations. The model will include mean RSA as the dependent variable. Predictors will include Epoch (baseline, social interaction task, or challenge task), Intervention Group, and the interaction of these terms. Random intercepts for participant ID will account for repeated measures across epochs.

After fitting the model, hypotheses will be tested with a series of planned contrasts. Specifically, we will evaluate the statistical significance of the following contrasts:

- TP vs. Control at Baseline, Social Interaction, and Challenge (Hypothesis a; 3 tests)
- PN vs. Control at Baseline, Social Interaction, and Challenge (Hypothesis b; 3 tests)
- Effect of Social Interaction epoch with reference to Baseline for TP vs Control (Hypothesis c, 1 test)
- Effect of Social interaction epoch with reference to Baseline for PN vs. Control (Hypothesis d, 1 test)

- Effect of Challenge task epoch with reference to Baseline for TP vs Control (Hypothesis e, 1 test)
- Effect of Challenge task epoch with reference to Baseline for PN vs. Control (Hypothesis f, 1 test)

As a robustness check, we will also fit alternative models to test Hypotheses c and d. These will be linear regression models in which the outcome represents RSA withdrawal from baseline (Baseline RSA – Social Interaction RSA), and the predictor is Intervention Group. This model will not include Epoch or any random effects.

To address research questions related to the CIB, we will compare the three treatment groups along all sub-constructs of the CIB, with primary focus on comparing each intervention group to the control group. The CIB sub-constructs include three maternal outcomes (maternal intrusiveness, maternal sensitivity, and limit setting), three child outcomes (engagement, compliance, and withdrawal), and two dyad-level outcomes (dyadic reciprocity and negative states). We specifically hypothesize that maternal intrusiveness, child withdrawal, and dyadic negative states will be lower in both intervention groups compared to the control group, and that maternal sensitivity and limit setting, child involvement and compliance, and dyadic reciprocity will be higher in both intervention groups compared to control. We will fit linear mixed effects models for each outcome. The predictors in the models will include fixed effects for Epoch (social interaction task or challenge task), Intervention Group (reference=control group), and the interaction of these terms. Random intercepts for participant ID will account for repeated measures across epochs. Because we have no specific hypothesis that the intervention effect will vary across tasks, we will drop the interaction term if it is not significant based on a likelihood ratio test (LRT). In this case, we would interpret the main effects of each intervention group indicator using a LRT to test the study hypotheses.