Cover Page:

Statistical Analysis Plan

Title of the study:

A Risky Sex Prevention Intervention for Middle School Age Minority Girls

NCT number:

NCT02103218

Date of the document:

January 27, 2019

Statistical Analysis Plan

The study was powered to detect a 0.24 proportion difference in prevalence of risky sexual behavior with 80% power assuming a 0.05 significance level when the sample size was 40 girls per group at study's end. Also, the study was powered to detect an increase of 9 points on HIV knowledge, an increase of 3 points on self-esteem, an increase of 0.26 in racial pride, and an increase of 0.52 in assertiveness under similar assumptions.

Baseline characteristics were compared between intervention groups using *t*-tests or Wilcoxon rank-sum tests for continuous characteristics and Chi-square or Fisher's exact tests for categorical measures. Analyses were performed separately for girls and mothers. Attrition rates across study follow-up were compared between groups using Chi-square tests. Multiple imputation for missing data was performed in sensitivity analyses with 20 imputations using predictive mean matching with a fully conditional specification method (Berglund, 2014).

Inverse probability weighting (IPW) using propensity scores were used to account for the quasi-experimental study design (Austin, 2015; Stuart, 2007). The goal of such weighting is to balance the distribution of variables that would approximate balance as if a randomized controlled trial had been performed. Propensity scores were first estimated using pre-intervention baseline measures related to outcomes, as recommended by Brookhart (2006) and current literature. Standardized differences were compared before and after weighting for assessing whether adequate balance had been achieved (Austin & Stuart, 2015) as well as sensitivity analyses incorporating IPW and without were performed. Robust sandwich standard errors were additionally specified to account for propensity score estimation (Joffe, 2012).

Longitudinal comparisons of change over time between groups were performed using generalized estimating equations modeling (GEE; Fitzmaurice, Laird, & Ware, 2011). Outcomes

exhibiting skewness or suggestion of non-Normality were modeled using negative binomial count GEE. Model-based time-point specific differences were interpreted if omnibus tests for any differences in change over time between groups (i.e., group by time interaction) were significant (Fitzmaurice et al., 2011). A two-sided p-value < 0.05 was considered statistically significant.