ClinicalTrials.gov Data Entry Cover Sheet

Protocol ID: 22811 **Clinical Trials.gov ID**: NCT05988398

Project Title: Preventing Commercial Sexual Exploitation of Children (Project LIVE)

Principal Investigator: Katie M. Edwards, PhD, University of Nebraska-Lincoln

Statistical Design and Power

Power Analyses

Aim 1a. N/A

Aim 1b. There is no "magic bullet" for sample size in qualitative research.[127-129] The sample size proposed for focus groups is consistent with other research conducted by the PI and Co-I (Dalla) that should allow for saturation (i.e., the collection of new data does not produce additional insights).[130] However, the sample size may be adjusted until data saturation is achieved.[130]

Aim 1c. N/A

Aim 1d. Power analyses to support the sample size for the measurement development study (for multi-item measures) are based on prior simulation work. [131]The proposed sample size of *n*=878 baseline surveys from youth will be utilized by splitting the sample in half to conduct exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), each with 439 youth. A sample size of 439 youth is adequate for implementing factor analyses using structural equation modeling approaches and provides power (> .80) to estimate two factor models, five manifest indicators each, nonnormal data, and factor loadings = .80.

Aim 1e. Regarding qualitative data, see narrative above for Aim 1b. Also, power analyses were not conducted for descriptive data (e.g., frequencies of the number of sessions attended, extent to which sessions were delivered with fidelity). Regarding student survey data collected as part of the OPT, we will examine changes in intermediary outcomes (see Figure 1) over time given that there will likely not be a sufficient window of time to see changes in behavioral primary and secondary outcomes at the immediate post-test which will happen within a few days to a week or two following the ending of the programming. For the exploratory program impact analysis on intermediary outcomes (Figure 1), a sample size of n=878 (approximately .5 assignment rate to condition), will provide adequate power (.80; a=.05) to detect a minimum effect size (MDES) *d*=.20 as determined by G-Power analysis of covariance (ANCOVA) routine.

Aim 1f. N/A

Aim 2a. Power analyses to support the sample size for the effectiveness analyses for our primary, secondary, and intermediary outcomes (see Figure 1) were conducted using G-Power 3.1. Power was set to .80 and a=.05. Component A of the study will provide incidence rates which will be used to recalculate power prior to the implementation of Component B. For the purposes of this proposal, we calculated the MDES based on the proposed sample size. For the primary outcome of CSEC perpetration, there are no reliable incidence estimates in the literature on which to draw. With a sample size of 7,241 students enrolled, assuming conservative minimum incidence rates of CSEC perpetration of 1% to 5% (although we hypothesize it will be higher), we have power to detect an MDES difference between conditions, odds ratio (OR)=.49 to .75, respectively. Similarly, for secondary outcomes of CSEC victimization, we estimated a range of incidence rates from a conservation 5% to 50%+ based on literature discussed following for other types of victimization[96, 132, 133] as well as rates of CSEC youth at increased risk (e.g., system involved, homeless)[18-20], which will be present in our sample based on DMPS data, we have the power to detect an MDES difference between conditions, odds ratio (OR)=.75 to .88, respectively. For secondary outcomes of TDV/SV perpetration and victimization, given past 12- months incidence rates[132] (from nationally representative data from youth) of 63% for perpetration (of any type of TDV) and 69% for victimization (of any type of TDV), we have power to detect an MDES difference between conditions, odds ratio (OR)=.88 to .89. For sub-types of TDV (past 12-month incidence rates: physical victimization: 18%; sexual victimization: 18%; psychological victimization: 66%; physical perpetration: 12%; sexual perpetration: 12%; psychological perpetration: 62%)[132], we have power to detect an MDES difference between conditions, odds ratio (OR)=.83 to .89, respectively. For the secondary outcome CSEC bystander action and opportunity and intermediary outcomes (i.e., knowledge, valuing of self and others, healthy relationship skills, bystander efficacy, perceptions of school social norms, prevention conversations with school personnel), the ANCOVA routine suggested power (.80) to detect a small difference between groups (MDES *d*=.06).

Aim 2b. For the mediation analyses, based on simulation literature[134, 135], a sample size of 7,241 provides power (> .80) to detect small effects (*a*=.28; *b*=.14), which are typical of mediation designs. For the moderation analyses, we used simulation-based approach to power analysis using Mplus[131]. We used a multiple group model, assuming categorical moderators (e.g., gender). The proposed sample size provides sufficient power (> .80) to detect moderation group differences in treatment effects, *b*=.20. As an alternative way to examine power for moderation of treatment effects, we examined the MDES within moderation groups. For this analysis (G-Power ANCOVA), we took a conservative approach and used the moderator that may have the largest discrepancy in group size, such as gender identity and sexual orientation. Based on population data, we estimate 2.5% will identify as a gender minority (inclusively defined, e.g., trans, non-binary, gender queer, Two Spirit)[136-138], leaving a sample size of *n*=181; and 10% of the sample will identify as a sexual minority[113, 139], leaving a sample size of *n*=724. Simulation power analysis supports being able to detect a small condition group difference, MDES *d*=.20.

Aim 2c. N/A

Aim 2d. N/A

Aim 2f. N/A

Proposed Analyses

Aim 1a. N/A

Aim 1b. Under the leadership of Co-I Dalla, data will be analyzed using thematic analysis.[90, 91] More specifically, audio files will be transcribed (using Rev, a speech to text on-line transcriptions service), de-identified by a member of the research team, and then read thoroughly to become familiar with the data. Initial codes are then created that represent the meaning and patterns evident in the data. A codebook is created simultaneously to keep track of the codes. The next step involves reading data again and applying codes to excerpts/passages that correspond to codes. Excerpts that represent the same meaning(s) receive the same code. Next, all excerpts with the same code are grouped together. The codes are then grouped into potential themes; with some themes having sub-themes. Themes are then reviewed and revised, ensuring that each theme has enough data to support it as a distinct concept. Themes are then arranged in narrative communicating results of analysis. Although school personnel data will be analyzed separately from student data, we will examine similarities and differences in themes across these populations. We will also examine variations in themes across subpopulations of students as well as subpopulations of school personnel. Themes will be shared with the RAB and their input on how to use this data to inform programming and research will be solicited.

Aim 1c. N/A

Aim 1d. Rigorous psychometric evidence for the newly developed multi-item measures will be evaluated using multiple indices (as most applicable to each measure) as guided by the American Educational Research Association standards[92]. First, using data from the baseline (pretest), we will randomly assign students to either Phase 1 (EFA; *n*=435) or Phase 2 (CFA; *n*=435) to determine evidence of internal measurement structure. In Phase 1, we will examine the factor structure by conducting EFA using the common factor model and full information maximum likelihood (FIML) estimation with oblique rotation within a structural equation modeling (SEM) framework[140] in Mplus[141]. This rigorous method retains all cases with the assumption that any missing responses are missing at random (MAR) to maintain power and improve estimation under conditions of missing data[142]. We will employ parallel analysis to determine the appropriate number of factors[140] (and then evaluate the fit of the model within SEM)[140]. After conducting the EFA, we will then replicate the measurement structure in a new sample of students using CFA within SEM with FIML. We will also examine internal consistency estimates of reliability of our measures. Next, we will examine evidence of convergent associations between the newly developed measures with existing, psychometrically sound measures (see Table 3). Structural regression models will be estimated including a latent and manifest measurement model and regression paths to determine associations between the newly developed multi-item and single-item measures and existing measures to determine if associations are in the theoretically derived hypothesized direction. Model fit will be determined using multiple indices, including chi-square statistic, root mean square error of approximation (RMSEA :s 0.08), the comparative fit index (CFI □0.90), and the standardized root mean square residual (SRMR :s 0.10)[143].

Aim 1e. Regarding qualitative data, see narrative above for Aim 1b. Descriptive analyses will be used to gauge things such as the average number of sessions attended, the extent to which sessions were delivered with fidelity, etc. Regarding pre- to post- changes in intermediary variables, we will use analysis of covariance, including the baseline measurement of the outcome as a covariate, the condition as the independent variable, and the post-test measure of the outcome as the dependent variable. The analysis will examine differences between conditions to determine preliminary program impact and effect sizes. Maximum likelihood estimation will be used to account for missing data in the analyses. Classroom will be used as a cluster variable to adjust the standard errors for nesting.

Aim 1f. N/A

Aim 2a and 2b.

Preliminary Analyses.

We will guard against differential attrition across experimental conditions. We will track and examine if attrition rates differ for the experimental groups. We expect similar rates of attrition across experimental groups. Furthermore, we will use data from all enrolled cases, who all have baseline data based on our design in our analyses to limit the bias introduced by missing data. Missing data assumed to be at least MAR will be dealt with as a function of the data analytic process through maximum likelihood estimation (ML), which makes use of all available data and does not require deletion of incomplete cases.

Outcome Analyses (Aim 2a).

We will test the hypotheses related to examining effectiveness of participation in the RTS in leading to reductions in CSEC perpetration (primary outcome), as well as reductions in CSEC victimization and TDV and SV victimization and perpetration and increases in bystander intervention in CSEC situations (secondary outcomes). To improve the rigor of analyses for Aim 2a, we will use propensity-score analysis to reduce the impact of selection bias, and latent constructs, when possible, to reduce measurement error, reduce paths and variables, and increase statistical power (MacCallum et al., 1996). This approach produces rigorous estimates of causal effects within the context of quasi-experimental designs when random assignment is not possible by controlling confounders, including selection effects[144]. This approach will use four analytic steps: (1) estimate propensity scores in a separate analysis, (2) use the propensity scores to adjust for confounding, (3) assess balance, and (4) estimate the propensity score-adjusted treatment effects (Lanza et al., 2013). In step 1, to estimate the propensity score, we will consider all pertinent confounders and covariates (see Measurement Table 3). This step will use logistic regression of condition on the potential confounders to obtain each student's propensity score. All measured confounders that are predictive of selection into the experimental conditions and the outcomes will be included in the propensity model. Standardized mean differences in covariate values across treatment and comparison conditions will be used to determine the final set of confounders to include in estimating the propensity score. In step 2, an initial treatment effect model is estimated that adjusts for confounding using the propensity score. We will use inverse probability weighting (IPTW) to adjust for confounding. The standardized mean differences between experimental groups are computed on the weighted sample included the IPTW-adjusted propensity score (Step 2). If IPTW does not provide optimal balance, then other methods will be considered (e.g., matching). Then Step 3 is used to determine balance on the potential confounders included in the propensity score across the experimental groups. The goal is to determine whether there are any differences remaining on the confounders between experimental groups (balance = standardized mean differences < |0.2| after adjustment). Confounders will be adjusted and the propensity score model (Step 1) will be re-estimated until balance is reached.
Once balance is determined, the outcome analysis will be conducted in Step 4 via SEM. The experimental effect of interest will be the Average Treatment Effect (ATE), which is the population-level average effect of condition on the primary and secondary outcomes as detailed in the logic model. The effects of interest quantified are differences between experimental conditions at post-test and follow-up on outcomes (for continuous measures), the propensity score as a covariate. These effects indicate whether the outcomes in the experimental condition differ significantly from those in the comparison condition. At post-test and follow-up, we expect the difference between the experimental conditions to be significantly greater than zero, reflecting better outcomes for those in the RTS program as compared to those in the wait-list condition, after adjusting for the

propensity score. For categorical variables, logistic regression will be used to examine differences in prevalence by condition. The experimental condition is the independent variable in these models. Multilevel modeling is not possible given the small number of schools; thus, school will be included as a covariate. Classroom will be used as a cluster variable to adjust the standard errors for nesting. Models for Step 4 will be evaluated for overall model fit via x2 statistic, RMSEA :s .08, SRMR :s .10, and CFI □ .90. We will evaluate our hypotheses across several indicators, including significance (p), confidence intervals, and the standardized difference between groups (d), as well as the directionality of included path coefficients, and odds ratios.

Mediation and Moderation Analyses (Aim 2b).

We will assess mediators (e.g., perceptions of norms intolerant of CSEC) and demographic moderators (e.g., gender) of program effectiveness, including the propensity score and school as covariates in all models. To test for mediation, we used the product of coefficients method using bias-corrected bootstrapping with 1000 resamples to calculate the confidence intervals[145] (MacKinnon, 2012). For our moderators (e.g., age, gender, racial groups, sexual orientation), we will estimate structural equation multiple group models, first allowing estimation of the path coefficients to vary freely across groups, and then constraining paths to be equal when a structural path coefficient of interest was significant for one group and not the other group. We will conduct model comparisons using the chi-square (x2) difference test (i.e., x2, $p < .05$ indicating moderation).

Aim 2c. Descriptive statistics will be used to examine the extent to which sessions were delivered with fidelity. Process observations will be analyzed using thematic analysis as described above for Aim 1b.

Aim 2d. Descriptive statistics will be used to examine the frequency of responses to close-ended questions (e.g., To what extend did you like the session today?). Open-ended questions will be analyzed using thematic analysis as described above for Aim 1b.

Aim 2e. Interview data will be analyzed using thematic analysis as described above for Aim 1b.

Aim 2f. The purpose of this aim is to determine costs associated with the program implementation to inform future economic evaluation of the RTS. The process for documenting costs are described in Section C.18.