Study Title

Culturally Modified Family Based Therapy for Haitian Youth in South Florida (CIFFTA).

NCT number: 03876171

Date: July 19, 2018

Major Analytic Methods Proposed (Including Statistical Design and Power)

Specific Aims

Aim 1. To adapt CIFFTA and the assessment instruments necessary to test its efficacy to the unique Haitian sociocultural contexts. This process of adaptation has the purpose of optimizing CIFFTA's acceptability and cultural appropriateness. It will rely on ethnographic methods (participant observations, focus groups, and interviews) to guide the adaptation process.

Aim 2. Determine, via a pilot efficacy trial, the impact of CIFFTA compared with standard-of-care for reduction of drug use, risky sexual behaviors and delinquency risk among Haitian adolescents ages 13 to 17 in Miami-Dade County. The ethnographic component of the study will monitor the delivery of the intervention, collecting process data and affording the investigators needed perspective on the intervention's impact.

Aim 3. To generate preliminary data on the usefulness of assessment tools and CIFFTA's efficacy among the population. These data will help evaluate the effect of the intervention and its best form of implementation. They will also provide preliminary evidence to justify a full clinical trial of CIFFTA in a population of Haitian youth at risk.

Analytic Methods

Latent Growth Modeling within a Structural Equation Model (SEM) framework. We will conduct latent growth curve analyses using Stata 14 gsem (1), SAS Proc Mixed (2), SPSS Mixed Models, and/or Mplus 8 (3) to compare the trajectories of participants in CIFFTA to those in TAU. Given that the data are balanced and have more than 2 time points, the model is identified, thereby allowing us to perform SEM growth curve analysis (4). These models conceptualize each component of growth (the baseline/intercept, the slope/rate of change, the quadratic slope, etc.) as composed of a fixed and random effect. Thus, these models have multiple components of variance as well as a regression-type model of the underlying mean growth curves of the two groups. Within an SEM framework, growth curves are treated as latent (unobserved) variables influenced by occasion- and individual-level covariates, allowing us to explicitly model growth trajectories overtime.

Specific analysis link functions will be matched to fit the distribution of each outcome variable. For ordinal outcomes (e.g., frequency of drug/alcohol usage), we will use an ordinal logistic regression function. The proportional odds assumption will be evaluated using the Brant test, and if needed, addressed by allowing problematic coefficients to vary across levels of the outcome variable. Count outcome measures (e.g., number of sexual partners, problem behaviors, re-arrests) will be analyzed using a negative binomial function rather than Poisson regression to address potential over-dispersion. Vuong tests will be performed to assess whether the data have an excess of zeros, and if necessary, zero inflated negative binomial regression will be used instead.

Analysis for Aims 2 and 3: The first set of analyses will explore the efficacy of CIFFTA compared to SOC in modifying the main adolescent outcomes and to generate effect sizes for a possible full clinical trial: a) drug/alcohol use, b) risky sexual behavior, c) delinquent behavior, and d) recidivism. Latent growth modeling will be used to investigate the efficacy of CIFFTA as compared to SOC. For each model, we will examine the effect sizes to determine the overall impact of CIFFTA over SOC on the outcome variables. We will generate estimates of effect sizes, such as Cohen's d (5), for both rate of change and for the

difference at a particular time between the two conditions (6). Following Cohen's guidelines (5), we will interpret our effect sizes using the convention of "small" (d = .2), "medium" (d = .5), and "large" (d = .8).

Our hypotheses concerning the direct and indirect effects of key variables on drugs/alcohol use, risky sexual behavior, delinquent behavior, and recidivism will be tested using SEM mediation analyses. To test potential indirect relationships we will include several mediators, including family connectedness, acculturation (bicultural involvement scale**)**, and motivation to change. This will allow us to examine whether CIFFTA directly effects our outcomes of interest and/or whether the effect of CIFFTA operates through its effect on family connectedness, acculturation, and motivation to change. These mediation models will enable us to address **Specific Aim #2** by testing the significance of paths in the model linking the intervention variable to the risk-related outcomes in question (both at discrete points in time and as the change index derived from differences in ordinal frequency at baseline and at post-intervention time intervals).

In a separate set of analyses, we will examine the hazard rate of drug/alcohol use and sexual risk behavior during the post-intervention phase of the study. Comparisons across the two intervention conditions will thus be made regarding the hazard rate of having sex during the post-intervention period, sex with multiple partners (>1 per month), and unprotected sexual episodes. Specifically, shared frailty Cox proportional hazard regression will be utilized to examine possible differences in the hazard rate of various risky behaviors (drug- and sex-related) in CIFFTA and control conditions during the post-intervention phase. This will prove particularly helpful in determining the hazard rate for post-intervention sexual initiation or unprotected sex in the two groups being compared, while also controlling for the time-dependence nature of our data. Similar analyses will be performed regarding initiation into drug use and other problem behaviors to determine whether the possible benefits of the CIFFTA approach extend to a range of risky behaviors.

Power Analysis: We will recruit 88 participants for this study, with the intention of analyzing data for 80 participants after attrition. The original CIFFTA randomized trial with 28 Hispanic adolescents showed CIFFTA to be more efficacious than the comparison family treatment at: 1) reducing drug use, and 2) improving parenting practices as reported by adolescents (7). More recently, a technology assisted CIFFTA study of 80 minority youth at high risk for drug use showed CIFFTA to be more efficacious in reducing behavior and conduct problems and in improving family cohesion, when compared to a delay treatment condition (8). Based on these findings, CIFFTA was placed on the National Registry of Evidence Based Practices using their more rigorous criteria for entry. Given such consistent findings of efficacy for CIFFTA among Hispanic youth, a sample size of 80 will be sufficiently powerful to detect relatively small effects (R-square = 0.05) at adequate levels of power (0.84 - 0.86) using a relatively simple SEM. A sample size of 80 is also well above sample guidelines for Stage 1 treatment trials (80)). Nevertheless, to deal with attrition, which in the case of Haitian and because of the ethnographer's engagement with the families will not be higher than 10%, we plan a sample of 88 adolescents.

Nevertheless, we recognize and appreciate the limits of estimating simple SEMs, such as ours, with a small sample size. Therefore, we will take several steps to mitigate concerns about statistical power, including: limiting the use of latent variables when possible, constructing parsimonious path models, and using Bayesian estimators. When theoretically and empirically appropriate, we will use principal components factor analysis to combine multiple indicators into a single scale capturing the underlying construct rather than estimating multiple latent variables since the latter often requires larger samples.

The use of such factor-score scales will enable us to conduct simpler pathway analyses with fewer predictors, thereby freeing up additional degrees of freedom. Our path models will be as parsimonious as possible, focusing on key theoretically grounded mediating relationships instead of engaging in exploratory data analysis with numerous pathways. When appropriate, path analyses may be performed separately to increase statistical power. Finally, we will employ Bayesian estimators when possible since they often perform better with small sample sizes (9). Despite these additional steps, we also emphasize that our analyses are preliminary in nature, primarily designed to show initial efficacy in anticipation for the submission of a RO1 grant designed to examine these relationships using a larger sample in a possible full clinical trial (see **Specific Aim # 3**).

Analysis Rationale: While there are numerous approaches to analyzing longitudinal data, including hierarchical linear models (HLM), we opted for a SEM approach since it provides greater modelling flexibility and better enables us to test our underlying theoretical model. In contrast to HLM, SEM allows for the simultaneous estimation of multiple growth curves, indirect/direct effects, and multiple indicators of model fit. Mediating relationships can be assessed within a HLM framework using Baron and Kenny's (10) approach, but SEM allows for the decomposition of direct/indirect effects as well as the simultaneous estimation of mediating and moderating relationships (11). Moreover, SEM will allow us to estimate growth-curves and mediating relationships at the same time in a way that HLMs cannot (12). The estimation of indirect effects within a growth-curve framework is especially important for **Specific Aim #2**, as we hypothesize that CIFFTA will not only have a direct effect on our outcome measures, but will also have an indirect effect operating through its effect on family connectedness, acculturation, and motivation to change. The estimation of mediating relationships will provide greater insights into any causal mechanisms (if any) at work, allowing for a more refined application of CIFFTA in a possible full clinical trial. SEM is also better equipped for assessing our theoretical model. In addition to providing measures of statistical significance, SEM generates several useful measures of overall model fit (e.g., CFI, AIC, BIC, etc.) that can be used to evaluate the appropriateness of our theoretical model over alternative models (11). Finally, SEM is better equipped to handle missing data than traditional regression or HLMs as it allows for maximum likelihood or Bayesian estimation (13).

Survival analyses can also be performed within a SEM framework. We will utilize shared frailty proportional hazard Cox regression as it does not make assumptions about the baseline hazard rate like other survival analyses (e.g., exponential, Weibull), while the shared frailty component of the model will allow us to adjust for the dependence of survival times via random effects (12). Survival analysis is ideal for estimating time-to-event data such as the hazard rate of drug/alcohol use and sexual risk behavior during the post-intervention phase. Because our study contains 4 waves, each with their own start/finish date, the data will be right-censored, thereby necessitating the use of survival analysis techniques. While ordinary least squares regression treats censored cases as missing, survival analysis will adjust for censoring when calculating hazard rates, providing more accurate estimates (12).

References

1. StataCorp, L.P. (2015). STATA 12 [Computer software]. College Station, TX: StataCorp LP.
2. Littell, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. & Schabenberger, O. (1996). SAS for mixed models. 2nd Ed. Cary, NC : SAS Institute.
3. Muthén, L.K. & Muthén, B.O. (2007). Mplus user's guide. 5th Ed. Los Angeles, CA : Muthén & Muthén.
4. Rabe-Hesketh, S. & Skrondal, A. (2013). Multilevel and longitudinal modeling using Stata. STATA press. 3rd edition.
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale: Lawrence Erlbaum Associates.
6. Raudenbush, S.W. & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of treatment effects on polynomial change. *Psychological Methods.* 6 : 387-401
7. Santisteban, D. A., Mena, M.P. & McCabe, B.E. (2011). Preliminary results for an adaptive family treatment for drug abuse in Hispanic youth. *Journal of Family Psychology*. 25(4): 610-614.
8. Santisteban, D.A., Czaja, S., Nair, S., Mena, M.P. & Tulloch, A. (2017). Computer Informed and Flexible Family-Based Treatment for Adolescents: A Randomized Clinical Trial for at-Risk Racial/Ethnic Minority Adolescents. Behavior Therapy, 48(4), 474-489.
9. van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & Van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. European Journal of Psychotraumatology, 6.
10. Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology. 51*, 1173-1182.
11. Kline, R.B. (2015). Principles and practice of structural equation modeling. Guilford publications.
12. Dévieux, J.G., Jean-Gilles, M., Frankel, A., Attonito, J., Saxena, A., and Rosenberg, R. (2016). Predictors of sexual activity in Haitian- American adolescents. Journal of Immigrant Minority Health. 18, 161-172.
13. Kaplan, D. & Depaoli, S. (2012). Bayesian Structural Equation Modeling. In Handbook of Structural Equation Modeling. Edited by Rick H. Hoyle. Guilford Press.