Study Title:
Social Ecological Based Smoking Cessation Intervention in Public Housing Neighborhoods

 # R01HL090951

Date: 8/09/2018

<u>Statistical Analyses</u>.

<u>Preliminary Analysis.</u> Univariate descriptive statistics and frequency distributions will be calculated as appropriate for all variables. These analyses will identify departures from distributional assumptions for proposed procedures. If departures are identified, appropriate transformations of data will be applied, or alternative analysis procedures (e.g., semi-parametric or nonparametric) will be employed. Baseline values for demographic and other putative prognostic variables will be compared for imbalance across the two treatment conditions (TX, C). These analyses will identify potential confounding variables to be used as covariates in subsequent analyses. Putative prognostic variables that will be investigated through these descriptive analyses include age, income, education, nicotine dependence score, and depression score. *To investigate potential limits on generalizability, we will compare demographic and other baseline characteristics of the dropouts with those who were completers.*

<u>Premature exits (drop-outs) and missing data</u>. *In order to minimize potential bias due to missing observations, we will carefully document why data are missing, which will involve contacting via telephone, if necessary, those who do not return for follow-up assessments to attempt to ascertain smoking status, and reason(s) for dropout We will adopt several approaches to dealing with missing data for the ITT analysis set. For the primary dichotomous cessation outcomes, we will impute outcome status in two ways: (1) by considering dropouts as smokers, and (2) using a stochastic regression imputation developed by Hedeker [143] taking different source of variation into account. For analyses of continuous variables using a single outcome (e.g., change from baseline in nicotine dependence score), we will use the multiple imputation methods of Little [144] and Rubin [145]. For longitudinal data, where appropriate, we will employ longitudinal data methods (e.g., mixed models, random regression, or hierarchical linear models) which allow for missing data [143,146]. We will carry out sensitivity analyses by comparing the results of analyses using different methods of imputation to evaluate the role of missing data in the study's conclusions. The implications of the sensitivity analyses will be discussed in presentations of study findings. In addition, we will model the dichotomous outcome, missing/not missing end-of-study score, using logistic regression to determine if "missingness" is related to treatment group and/or initial level of the given outcome variable (baseline score). We will also use information on drop-outs as one of the outcome measures of effectiveness of the interventions being compared. We will make every attempt to keep subjects in the study and to obtain required measurements.*

<u>Analyses for Primary Aim</u>.

Aim #1. To test the effectiveness of a multi-level intervention (*Sister to Sister*) on smoking cessation outcomes in female smokers residing in public housing neighborhoods.

<u>Hypothesis 1.1</u>: As compared to the control group, women receiving the *Sister to Sister Intervention* will have higher 7-day point prevalence quit rates at 6-and 12-months.

<u>Hypothesis 1.2</u>: As compared to the control group, women receiving the *Sister to Sister Intervention* will have higher 6- and 12-month prolonged abstinence.

To evaluate the relationships between intervention outcomes and intervention status (TX or C), we will use a generalized linear mixed models (GLMM) approach *(or equivalently hierarchical linear models [HLM], random regression models [RRM], or mixed effects models [MEM]) with treatment assignment as the primary independent variable and measures of* *effectiveness as the dependent variable [146]. This approach accommodates a wide range of distributional assumptions for continuous and categorical outcome variables, including binary (e.g., smoking cessation/no cessation - the primary outcome), count, and ordinal measures, as well as multilevel data such as longitudinal measurements on subjects and a possible cluster effect within neighborhoods [143,146].These analyses allow for measurement of subjects at different time points, missing data, and time varying or invariant covariates.*

In the GLMM analyses, treatment (TX vs C) and measurement time (6 months and 12 months) will be included as a fixed effect, a term representing neighborhood (the cluster effect) will be included in the model as a random effect in order to account for correlation among women within the same neighborhood*, with a pair-specific fixed intercept parameter to represent each neighborhood pair [148].* The statistical test of primary interest will be the F-test for the treatment x time interaction, which will reflect a difference in change between the intervention and control groups over the course of the study.

In further analyses, putative covariates will be added to the model to examine whether the intervention remained statistically and clinically associated with cessation after controlling for the potential covariates. Each covariate considered for inclusion will be examined individually for a relationship with cessation outcomes. In the second step, those variables with a p value < .25 will be included in an initial model. Next, the potential confounder variable in the initial model with highest p value will be removed and the model will be run again. If the removal of the potential confounder variable does not result in a significant reduction in model fit (as indicated by a change in the model -2 log likelihood), then the variable is removed from further steps. The removal and subsequent testing of change in model fit will be repeated until all non-significant potential confounders are tested. Potential covariates include demographic variables (i.e., age, income, education), nicotine dependence score, depression score, body mass index, perceived stress, social smoking influences, *social ties, neighborhood cohesion, neighborhood smoking prevalence, neighborhood stress,* and time dependent covariates of self efficacy, spiritual well-being, social support, and *cessation resources.*

The magnitude of intervention effect sizes (e.g., differences in 7-day point prevalence proportions between TX and C groups, and odds ratios for cessation) will be estimated using 95% confidence intervals (95%CI). Effect size estimates allow evaluation of the "clinical" or practical significance/relevance of the study findings.

<u>Analyses for Secondary Aims.</u>

Aim #2. To evaluate the effect of the *Sister-to-Sister* intervention on individual (self-efficacy, spiritual well-being), interpersonal (social support), and neighborhood (cessation resources) factors and to determine if these factors mediate the effect of treatment on smoking cessation outcomes.

<u>Hypothesis 2.1</u>: As compared to the control group, women receiving the *Sister to Sister* intervention will have greater improvement in individual factors (self-efficacy, spiritual well-being), interpersonal factors (social support), and neighborhood factors (*cessation resources*) at 6-and 12 months.

<u>Hypothesis 2.2</u>: Positive changes in individual factors (self-efficacy, spiritual well-being), interpersonal factors (social support, and neighborhood factor (*cessation resources*) will mediate the effect of the 6- and 12-month cessation outcomes in women receiving the *Sister to Sister* intervention.

A generalized linear mixed models (GLMM) approach, as described for the Primary Aim, will be used for Aim 2. To examine the effect of treatment (TX vs C) on individual factors (self-efficacy, spiritual well-being), interpersonal factors (social support), *and neighborhood factors (cessation resources)* [*Hypothesis 2.1*], each of the individual variables within the factors will be used individually as the outcome (dependent) variables in the GLMM. As described above, treatment and measurement time will be considered fixed effects, and neighborhood will be considered a random effect, with a *pair-specific fixed intercept parameter to represent each neighborhood pair. Additional covariates, as described above, will be added to the model to determine the effect of treatment on these outcomes adjusted for the effect of the covariates.*

The analyses for Hypothesis 2.2 address the question of whether the individual (self-efficacy, spiritual well-being), interpersonal (social support), and neighborhood factors (cessation resources) explain (mediate) the relationship between the outcome (6- and 12-month cessation) and the intervention (Sister-to-Sister). In mediator analyses, it is posited that variation in a given independent variable (intervention) accounts for variation in the mediators (self-efficacy, spiritual well-being, social support, cessation resources) and variations in mediators account for variations in outcome (6-and 12-month cessation). For these analyses, we will follow the methods of Baron and Kenney [149], Holmbeck [150], and Kraemer [151]. GLMM analyses as outlined above will evaluate the relationship between the outcome (6- and 12-month cessation) and the intervention (Aim 1). Next, the relationship between intervention and the putative mediator variables (self-efficacy, spiritual well-being, social support, cessation resources) will be evaluated as described for Hypothesis 2.1. Finally, the mediating effect of self-efficacy, spiritual well-being, social support, cessation resources on a significant treatment effect will be evaluated in a GLMM model that simultaneously regresses treatment (TX vs C) and the putative mediating variables on outcome (smoking cessation). A significant smoking cessation/treatment relationship that becomes nonsignificant following the addition of the putative mediator covariable suggests that the latter accounts for (mediates) the former observed relationship. For example, we conclude that the relationship between treatment and smoking cessation is mediated by self efficacy if the variable representing self efficacy is significant and a previously significant treatment effect becomes nonsignificant in the multivariable model containing both.

143. Hedeker D, Mermelstein RJ, Demirtas H. Analysis of binary outcomes with missing data: missing=smoking, last observation carried forward, and a little multiple imputation. Addiction. 2007.
102,10:1513-1684.
144. Little RJA, Rubin DB (1987*)* Statistical Analysis with Missing Data*.* 1987. J. Wiley & Sons, New York.
145. Little RJA, Rubin DB (1987*)* Statistical Analysis with Missing Data*.* 1987. J. Wiley & Sons, New York.
146. Gibbons R. Longitudinal Data Analysis. 2006. John Wiley and Sons, Inc. New York.

147. McCulloch CE, Searle SR. Generalized, Linear, and Mixed Models. 2001. John Wiley & Sons, Inc, New York.
148. Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based
inference for community intervention trials. Statistics in Medicine. 1996.5,11:1069-1092.
149. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research:
Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology.
1986;51:1173–1182.
150. Holmbeck, GN. Toward terminological, conceptual, and statistical clarity in the study of mediators and
moderators: Examples from the child-clinical and pediatric psychology literatures. Journal of Consulting
and Clinical Psychology. 1997;65(4):599-610.
151. Kraemer HC, Wilson GT, Fairburn CG, & Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry,* 2002;*59:*877-883.