**RAISE ETP STATISTICAL ANALYSIS PLAN**

**Updated, October 14, 2014. Final Version**

**Cluster Randomization**

The RAISE-ETP study employed cluster randomization. After site selection, sites were randomized to deliver either NAVIGATE (the experimental intervention) or Community Care (treatment as usual, i.e. prevailing treatment at the site.) This choice, rather than randomization of individuals within sites, was based on the following considerations.

Randomization by site does not assume that sites are identified from the population of all sites in the US but it does mean that the method of identification will allow generalization beyond those sites that participate in the trial. In preparation of the original contract proposal, we approached a number of service delivery systems and ascertained that these system administrators believed sites approached for this study would agree to randomization by site, i.e., all patients who sign consent at a site that is randomized to the Intervention will receive the RAISE Intervention and all those patients who sign consent at a site that is randomized to usual care will receive usual care. This was validated by the sites that were included in the study. Following disclosure of randomization assignment to the sites, none withdrew consent to participate. There are several other advantages to this approach. Considering the relatively low rate at which new cases of first episode psychosis present for treatment at any given community mental health center, randomization by patient would result in a significant challenge to centers of creating and sustaining an enhanced intervention team of clinicians, since only one-half of the potentially eligible patients would be randomized to receive the Intervention. The consent process is also greatly simplified in a site randomization because patients do not have to agree to randomization, they only have to agree to participate in a study where they will receive the treatment that will be provided in their setting, as well as to understanding to which treatment condition the site has been randomized and that there are two conditions. With site randomization there will be no potential contamination or spill-over of experimental and control interventions within a site because all patients at that site will receive either one intervention or the other, thus limiting clinician and patient exposure to only one intervention within each site.

The disadvantages of site randomization include the risk that the matching of sites will not be effective in minimizing imbalances between sites and that there will be systematic differences between intervention and usual care sites, i.e., that cluster randomization will not be successful. However, as discussed below, it is possible to make adjustments in the statistical analysis for imbalances involving observed covariates.

**Statistical Implications of Randomization by Site**

The statistical considerations and methods to be employed include the randomization method and addressing issues involving analysis of data at the participant level while randomizing at the site level. First, the randomization of the 34 sites entailed a randomization of sites within groups. The matching of sites was based on three factors that are thought to be highly related to the QLS outcome: race, ethnic distribution of treated patients and position on an urban, suburban, rural continuum. The sites were sorted on the basis of their assigned values for each of these factors. Such a strategy has been used for other recent psychiatric clinical trials in which clinical sites have been randomized (e.g., Bruce et al. 2004). Although site is the unit of randomization, the participant is the unit of analysis. Following many precedents in

the statistical and medical literature (e.g., Murray et al. 2004; Bruce et al. 2004), we propose to use random effects longitudinal models to perform intent-to-treat comparisons of the two randomization treatment groups at each of the follow-up visits. Other approaches (e.g., permutation tests of site-level proportions and means; generalized estimating equations estimation for models without random effects (GEE) (Diggle et al., 2002)) have been proposed for such randomization designs. However, these approaches make more restrictive missing data assumptions than the proposed random effects approach. Moreover, the permutation approaches cannot easily adjust for potential observed confounders, nor can it pool data from all follow-up visits to improve power in assessing differences at specific visits, e.g., 12 and 24 months. Consequently, the random effects approach is the preferred analytic method in this context.

## Statistical Analysis

As indicated above, this study will use a cluster randomization strategy in which each clinical site is randomized to one treatment. Such a design, based on randomization of sites, but with collection of data at the patient level across time, allows an analysis at the patient level of randomized treatments (e.g., Donner 2004; Bruce et al. 2004; Small et al. 2006). Initially, baseline demographic and clinical characteristics will be examined. Frequency distributions will be produced. Measures of central tendency (mean, median) and variability (standard deviation, minimum and maximum) will be estimated on each continuous measure. Proportions will be estimated for categorical variables. Graphical displays (e.g., histograms and boxplots) will be produced. Transformations will be used when distributional assumptions are not fulfilled for inferential tests on a continuous measure. The treatment groups will be compared on baseline demographic and clinical variables using two-level mixed-effects linear or logistic regression analyses. The demographic variables include Age, Race (White/Black/Other), Ethnicity (Hispanic/Non-Hispanic), Gender (M/F), Marital status (Cur/Past/Never), Current Residence (4 categories), patient's highest education level, mother's highest education, father's highest education (4 categories each), Student Status (YN), Working Status (YN), Employed (Ever/Never), Type of Insurance (Private/Public/None), Referral source (Usual/Outreach). The clinical variables include SCID psychosis diagnoses (6 categories), SCID diagnoses of lifetime use of alcohol and cannabis (3 categories), Antipsychotic medication status at time of enrollment (YN), Duration of untreated psychosis (mean), Age at first psychotic symptoms (mean), Number of previous Hospitalizations for psychiatric illness (0/1/2/3+), Current treatment setting (4 categories), Insight (mean; PANSS G12), Stigma scale (mean), SF12 (means; Physical/Mental), Quality of Life Scale (means; Total Score + 5 subscales), PANSS (means; Total Score +5 subscales), Calgary Depression Scale (mean; Total Score), CGI Severity Scale (mean), Mental Health Inpatient (number of days last 30), Mental Health Outpatient Visits (number in past 30 days), and the Fagerstrom Smoking and Chewing Tobacco scores at baseline (means). We will additionally report adjusted means for continuous variables and proportions for discrete variables after accounting for site clustering effects. We will not expect any baseline variable with more than 10% individuals having missing values. Two-level mixed-effects linear or (multinomial) logistic regression analyses will be employed for comparing the baseline covariates between the two treatments. In a similar manner, dropouts and completers will be compared on baseline variables. Unless specified otherwise, each of the statistical tests described will use a two-tailed alpha-level of 0.05.

### *Analyses of Primary Hypothesis*

**Hypothesis 1:** Subjects treated at RAISE-ETP Intervention sites (NAVIGATE) will exhibit greater improvement in Quality of Life Scale (QLS) scores over the course of the two-year trial than will those treated at the sites that provided Community Care.

The analysis of treatment differences will examine the repeatedly assessed continuous outcome (Total QLS score) from the assessment visits during the first two years (baseline, 6, 12, 18, 24) with a three level nested random effects linear model (e.g., Ten Have et al. 1999; Bruce et al. 2004). This mixed-effects linear regression model will include a random intercept and slope over time at the patient-level and a random intercept at the site level and a fixed effect for treatment indicator, time, and the interaction between treatment and time. This treatment by time interaction will be used to determine if the rate of improvement in QLS varies between two intervention groups over the course of treatment. The decision rule for Hypothesis 1 calls for rejection of $H_0$ if the treatment by time interaction is statistically significant (using a two-tailed alpha-level of 0.05). If the rates of improvement in patients receiving the two treatments are different, further analysis will include use of a discrete version of the measure of time instead of treating time as linear. The discrete version of the time measure will allow us to investigate any non-linear trend as well as the specific time points at which the two treatments have different outcomes.  Four dummy variables for time representing 6-, 12-, 18 and 24-months will be used with the baseline time = 0 as the reference category.

The test for site heterogeneity between the two treatments will be evaluated by testing the site-treatment-time interaction. Specifically, we will assign the sites in the two treatment groups to two different distributions for their random effects at the site-level, and the random effects of time will also have two different distributions for their random effects at patient-level.  Among the sites with the same treatment assignment, we will use Akaike's Information Criteria (AIC) to assess whether the variance component (in our case is the variance of the random effects) is zero.  If any one of the two variance components is statistically significantly different from zero at alpha=0.05, we will conclude that there is a site heterogeneity within treatments.  We will compare whether control sites have more heterogeneity than the NAVIGATE sites by testing whether the two variance components of the random effects are equal using the likelihood ratio test as follows. Specifically, we let the variance of the NAVIGATE sites' random effects be equal to Vs, and let the variance in the random effects of the control sites equal to Vx + ex. We will then test whether ex = 0  using a likelihood ratio test conducted with a two-tailed alpha-level of 0.05. If it is statistically significant, then we will conclude that the heterogeneity among the sites is different between the two treatments.

Likelihood ratio tests will also be used to compare the model fit with that having a first-order autoregressive (AR1) covariance structure, as described by Hedeker and Gibbons, 2006 (pages 101-111). The residuals and the predicted random effects from the mixed model will be checked for model diagnosis. If the residuals are highly skewed, we will consider transformation of the QLS measure or use a mixed-effects Poisson model instead of a linear model. If the random effects are not normally distributed, we will use fixed effects for the sites instead of random effects. The fixed-effects will use more degree of freedom (df = 33) while the random effects will use only one degree of freedom.  Furthermore, if the QLS has bimodal or multiple modal distribution, we will use a growth mixture model (Muthen et al., 2002) instead of a mixed model.

If significant baseline group imbalance is detected on a particular variable (at the site- or individual-level), and that variable is correlated with the outcomes on the QLS at a level of .30 or higher, the variable will be included as a covariate in the primary analyses of outcome. Site-level variables will be treated as an individual level covariate by letting the individuals within the same

site have the same value of the site-level variable. If we have more that 10 significant differences between groups, we will conduct mixed model logistic regression on the group assignment, with the variables plus the random effects of sites to identify a more parsimonious, non-overlapping set of variables. If there are still more than 10, we will construct propensity scores for group assignment and control for them to evaluate the QLS outcome (Leon et al,, 2007a). The propensity score would be calculated at the individual level as the estimated probability of receiving NAVIGATE vs. community care which would be computed through a logistic regression model of the significant variables with a random effect for sites.

***Secondary Hypotheses:***

**Hypothesis 2a:** Subjects specifically diagnosed with schizophrenia spectrum disorder treated at NAVIGATE sites will exhibit greater improvement in the Quality of Life Scale (QLS) over the course of the two-year study than will those treated at the Community Care sites.

These analyses will include only subjects with confirmed schizophrenia, schizophreniform and schizoaffective disorders at the one-year SCID assessment and will be conducted to parallel those described in Hypotheses 1.   It is estimated that approximately 335 (83%) of the 404 patients randomized to treatment will have confirmed schizophrenia spectrum diagnoses and will be included in this analysis. Since the goal of RAISE-ETP is to specifically improve the treatment of schizophrenia, this analysis will examine the QLS in that patient group.

**Hypothesis 2b:** Subjects treated at NAVIGATE sites will exhibit greater likelihood of cross-sectional *remission* (defined according to PANSS criteria set out by Andreasen et al 2005) over the course of the trial than will those treated at the sites that provide usual care.

**Hypothesis 2c:** Subjects treated at NAVIGATE sites will exhibit greater likelihood of cross-sectional *recovery* according to criteria of Robinson et al. (2004) over the course of the trial than will those treated at the sites that provide usual care.

The secondary analyses (Hypotheses 2b and 2c) of treatment differences will examine repeatedly assessed binary recovery and remission status from all four assessment visits (baseline, 6, 12, 18, 24, months). Separate mixed-effects logistic regression models will be used to examine group differences in recovery and remission status. The models will each involve three level nested random effects with a random intercept and slope over time at the patient-level and a random intercept at the site level and a fixed effect for treatment, time, and the interaction between treatment and time. (e.g., Ten Have et al. 1999; Bruce et al. 2004). The treatment by time interaction will be used to determine if the recovery or remission status differs between the two intervention groups over the course of the two years of treatment. Treatment heterogeneity across site will be evaluated similarly as for the primary continuous outcomes. Nonlinear trend will also be studied using dichotomous measures of time similar to the methods described for the primary outcomes.

**Hypothesis 2d:** Subjects treated at NAVIGATE sites will have higher outpatient treatment costs, but will have reduced use of costly inpatient care and improved quality of life which together will be sufficient to justify the additional costs of the intervention.

***Analysis of Cost Effectiveness:*** Methods to be used for the statistical analysis of cost effectiveness are described in detail in another document.

***Strategies for Attrition:*** Clinical trials are vulnerable to attrition over time due to inefficacy, adverse effects, and other reasons and this can introduce bias and reduce power, precision and generalizability (Leon et al., 2007b). Although we will make every effort to prevent attrition, we acknowledge that some attrition is inevitable in complex effectiveness RCTs that involve both pharmacological and psychosocial interventions. We have used a variety of methods to manage assessment attrition including maintaining telephone contact and conducting assessments by phone if necessary, allowing and encouraging subjects to return after they have dropped out for a time. Our power analyses (below) assumed 10% attrition each successive 6-month assessment period. The proposed mixed-effects models for Hypotheses 1 will incorporate all available data from subjects, even those who drop out of treatment or miss assessment visits since the study model continues to assess all subjects regardless of whether they are in treatment or not. Mixed-effects models yield valid inferences assuming ignorable attrition or data missing at random (MAR, i.e., attrition that is accounted for by measures of covariates, including treatment, or the dependent variable that are measured prior to dropout (Laird 1988)). Two strategies will be used to examine the sensitivity of the assumption of ignorable attrition. First, a pattern mixture model (Little, 1993) will be used to examine response to treatment among participants with various dropout patterns. For Hypotheses 1, this will be implemented in the mixed-effects framework described by Hedeker and Gibbons (1997) in which subjects are classified by attrition pattern (e.g., early dropout, middle dropout, late dropout, completer). Separate analyses will be used to compare the magnitude of the treatment effect across subjects classified by attrition pattern. (Because of the differential Ns expected across attrition patterns, this sensitivity analysis will focus on between group effect sizes.) Second, we asked subjects to rate their "Intent to Complete" the trial at baseline and to rate their "Intent to Attend" the next assessment session and at each prior assessment session using a Likert rating scale (Leon et al. 2007b). This variable will be used as a covariate to account for attrition in an effort to fulfill the assumption of ignorable attrition. The estimates of the treatment effect from the models described above (Hypotheses 1, 2a – 2c) will be compared with models that also include the main effects and interaction of treatment and either dropout pattern or measures of Intent to Attend. These two approaches will be used to examine the sensitivity of the results to attrition.

If the dropouts are suspected of being not ignorable or MAR by the sensitivity analysis, then they are missing not at random (MNAR) which means that dropping out depends on unobserved factor(s). In this case we will use a joint modeling approach in which the two mixed-effects models for QLS and for dropping out will be linked by sharing a latent variable representing unobserved factors for dropout (Wu and Follmann 1999).

**Exploratory Analyses:** Exploratory analyses will examine several mediators and moderators of treatment. Moderators will include gender, ethnicity, age, duration of untreated psychosis (DUP), premorbid functioning, cognitive impairment, substance use and abuse. Mediators will be post-baseline factors: substance abuse, psychopathology measures assessed by the PANSS and CDSS and adherence to medication and psychosocial treatments. The following specific relationships will be assessed.

Moderators: It is hypothesized that there will be an added benefit of RAISE/NAVIGATE for those who at baseline characterized by: older age, shorter duration of untreated psychosis, the combination of better premorbid functioning and shorter duration of untreated psychosis, less cognitive impairment, or no substance use and abuse. Each of the hypothesized moderators meet the criteria for a moderator set forth by Kraemer et al. (2002). That is, a moderator must 1) precede treatment; 2) be uncorrelated with treatment; and 3) have interactive effect with the

treatment-visit slope interaction on outcome (moderator by treatment by time).  Each of the hypothesized moderators that we have chosen meet the first criterion and we expect, based on randomization, that they will each meet the second. We will not examine differential treatment effects (i.e., the moderating effect) unless the first two criteria are met. Our analyses will focus on the third criterion, the interactive effect, i.e. a differential treatment effect on the slope across time for those with and without the potentially moderating characteristic (e.g., substance use).

Mediators:  It is hypothesized that there will be an added benefit of NAVIGATE for those who post-baseline:  do not have substance use during treatment, improve more in terms of psychopathology and exhibit better adherence to medication and to psychosocial treatments.

   Each of the hypothesized moderators meet the criteria for a moderator set forth by Kraemer et al. (2002). That is, a moderator must 1) precede treatment; 2) be uncorrelated with treatment; and 3) have interactive effect with the treatment-visit slope interaction on outcome (moderator by treatment by time).  Each of the hypothesized moderators that we have chosen meet the first criterion and we expect, based on randomization, that they will each meet the second. We will not examine differential treatment effects (i.e., the moderating effect) unless the first two criteria are met. Our analyses will focus on the third criterion, the interactive effect, i.e. a differential treatment effect on the slope across time for those with and without the characteristic (e.g., substance use).

   Mediators identify potential mechanisms of treatment. Accordingly, it is hypothesized that the treatment effect of NAVIGATE on the slope of the outcome, QLS, will occur by reducing substance use during treatment, improving psychopathology and exhibiting better adherence to medication and psychosocial treatments.  Under the longitudinal data framework, these time-varying mediators will be lagged such that we will use the measure of a particular mediator obtained at the visit prior to the visit at which the outcome measure is used in the analysis. In contrast to our moderator analyses, either a clinically meaningful main effect of the mediator adjusting for the treatment or a mediator by treatment interaction would provide evidence of a mediator effect (Kraemer et al., 2002). We will perform a Sobel-like test using a "product of coefficients" approach for the mediation effect as well.  A three-way interaction of treatment by time by mediator on the outcome will be also examined in order to see whether the mediator would change the effect of the intervention on the outcome slope.  We will consider a mediator-based change in magnitude of more than 20% of the treatment*time effect to be meaningful.  For a categorical mediator, we will adopt the approach of Iacobucci (2012). For multiple mediators, we will follow the approach outlined by VanderWeele and colleagues (VanderWeele TJ and Vansteelandt, 2013; Valeri and VanderWeele 2013).

   Based on the recommendations of Kraemer et al. the analyses of both moderators and mediators of treatment will focus on the magnitude of the effect and not on significance testing (Kraemer et al., 2002). For instance, we will compare the magnitude of the treatment*time interactions under the random effects models across the levels of categorical moderators (e.g., substance use and abuse:  positive vs. negative symptoms). Continuous mediators will be examined in a fashion similar to our description above. That is, if inclusion of the term representing the moderator by treatment interaction results in change in the magnitude of more than 20% of the treatment*visit effect, we would consider it meaningful.

   Finally, we note that any results based on these mediation analyses may be confounded by unmeasured factors, and hence causality cannot be assumed. As a sensitivity analysis, we will implement the causal mediation approach of Ten Have et al. (2007), which relies on using

the randomized intervention as an instrumental variable to protect the mediation analyses against unmeasured confounding. Such an analysis however requires additional assumptions that will be tested using the methods specified in Ten Have et al. (2007).

If the results of moderator analyses indicate that some groups are exceptionally responsive to NAVIGATE, this could be used to guide inclusion criteria for a subsequent study. In contrast, if the analyses detect some groups that are not at all responsive to NAVIGATE, efforts to develop more appropriate interventions for those groups can be proposed. Alternatively, the results of the mediator analyses could guide refinements of the treatment approach used in NAVIGATE, enhancing the likelihood that patients will receive the beneficial mediators and targeting other changes to reduce the risk of exposure to mediating events that render a participant vulnerable to treatment failure.

## References:

Andreasen NC, Carpenter WT, Kane JM, Lasser RA, Marder SR, Weinberger DR (2005). Remission in Schizophrenia: Proposed Criteria and Rationale for Consensus. *American Journal of Psychiatry* 162: 441-449.

Bruce ML, Ten Have TR, Reynolds CF III,  Katz  II, Schulberg HC, Mulsant BH, Brown GK, McAvay GJ, Pearson JL and Alexopoulos GS (2004).  Reducing Suicidal Ideation and Depressive Symptoms in Depressed Older Primary Care Patients: A Randomized Trial.  *Journal of the American Medical Association*, 291: 1081–1091.

Diggle P, Heagerty P, Liang K-Y, and Zeger S (2002).  Analysis of Longitudinal Data.  Wiley. Second Edition.

Donner A, Taljaard M and Klar N (2007).  The merits of breaking th matches: a cautionary tale. Statistics in Medicine 26 (9): 2036-51.

Hedeker D, Gibbons RD. Longitudinal Data Analysis. Hoboken NJ:  John Wiley and Sons, 2006.

Hedeker D, Gibbons RD (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* 2:64-78.

Kraemer HC, Wilson GT, Fairburn CG, and Agras WS (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry* 59 (10):877-883.

Iacobucci D (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology* 22: 582-594.

Laird NM (1988). Missing data in longitudinal studies. *Statistics in Medicine* 7: 305-315.

Leon AC, Hedeker D and Teres JJ (2007a). Bias reduction in effectiveness analyses of longitudinal ordinal doses with a mixed-effects propensity adjustment. *Statistics in Medicine* 26 (1): 110-123.

Leon AC, Demirtas H & Hedeker D (2007b). Bias reduction with an adjustment for participants' intent to drop out of a randomized controlled clinical trial.  *Clinical Trials* 4, 540-547.

Little RJA (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88 (421): 125-134.

Murray DM, Varnell SP, Blitstein JL (2004). Design and analysis of group-randomized trials: a review of recent methodological developments". *American Journal of Public Health* 94 (3): 423–32.

Muthén B, Brown CH, Masyn K, Jo B, Khoo ST, Yang CC, Wang CP, Kellam SG, Carlin JB, Liao J (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* 3 (4): 459-475.

Robinson D, Woerner M G, Alvir J M, Bilder R, Goldman R, Geisler S, Koreen A, Sheitman B, Chakos M, Mayerhoff D, & Lieberman J A (1999). Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Archives of General Psychiatry* 56: 241-247.

Robinson DG, Woerner MG, McMeniman M, Mendelowitz A & Bilder, RM (2004). Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry* 161: 473-479.

Small D, TenHave TR, Joffe M, and Cheng J (2006). Random effects logistic models for analyzing efficacy of a longitudinal randomized treatment with non-adherence. *Statistics in Medicine* 25:1981-2007.

Ten Have TR, Kunselman A, and Tran L (1999). A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine* 18:947-960.

Ten Have T, Joffe M, Lynch K, Maisto S, Brown G, and Beck A (2007). Causal mediation analyses with rank preserving models. *Biometrics* 63: 926–934.

Wu MC and Follmann DA (1999). Use of Summary Measures to Adjust for Informative Missingness Repeated Measures Data with Random Effects. *Biometrics* 55: 75-84.

Valeri L and VanderWeele TJ (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods* 18:137-150.

VanderWeele TJ and Vansteelandt S (2013). Mediation analysis with multiple mediators. *Epidemiologic Methods*, 2:95-115.