



## NIH PUBLIC ACCESS

## Author Manuscript

*J Subst Abuse Treat.* Author manuscript; available in PMC 2013 June 12.

Published in final edited form as:

*J Subst Abuse Treat.* 2011 January ; 40(1): 3–17. doi:10.1016/j.jsat.2010.07.010.

## Estimating Statistical Power for Open Enrollment Group Treatment Trials

Antonio A. Morgan-Lopez, Ph.D.<sup>1</sup>, Lissette M. Saavedra, Ph.D.<sup>2</sup>, Denise A. Hien, Ph.D.<sup>3,4</sup>, and William Fals-Stewart, Ph.D.<sup>5,6</sup>

<sup>1</sup>L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599

<sup>2</sup>RTI International, Division of Health Services and Social Policy Research, Research Triangle Park, NC 27709

<sup>3</sup>Department of Psychology, City University of New York at City College, New York, NY, 11030

<sup>4</sup>College of Physicians and Surgeons, Columbia University, New York, NY, 11032

<sup>5</sup>University of Rochester, School of Nursing, Rochester, NY, 14642

### Abstract

Modeling turnover in group membership has been identified as a key barrier contributing to a disconnect between the manner in which behavioral treatment is conducted (open enrollment groups) and the designs of substance abuse treatment trials (closed enrollment groups, individual therapy). Latent class pattern mixture models (LCPMM) are an emerging tool for modeling data from open enrollment groups with membership turnover in recently proposed treatment trials. The current article illustrates an approach to conducting power analyses for open enrollment designs based on Monte Carlo simulation of LCPMM models using parameters derived from published data from an RCT comparing Seeking Safety to a Community Care condition for women presenting with comorbid PTSD and substance use disorders. The example addresses discrepancies between the analysis framework assumed in power analyses of many recently-proposed open enrollment trials and the proposed use of LCPMM for data analysis.

### Introduction

In the early part of this decade, the majority of federally-funded behavioral treatment trials were delivered using individual, one-on-one counseling formats while, in community settings, group therapeutic approaches were the predominant delivery modality for substance abuse and alcoholism treatment (NIDA, 2003; Weiss, Jaffe, de Menil & Cogley, 2004). In response to this disconnect, NIDA and NIAAA sponsored requests for applications (RFAs) specifically geared to support group therapy efficacy trials (e.g., RFA-DA-04-008; NIDA/NIAAA, 2003a) and amended existing program announcements (PAs) to focus on group-delivered therapies (e.g., PA-03-126; NIDA/NIAAA, 2003b); in particular, these

© 2010 Elsevier Inc. All rights reserved.

Address correspondence to: Antonio A. Morgan-Lopez, Ph.D., Associate Professor of Quantitative Psychology, L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, CB#3270 Davie Hall 356a, Chapel Hill, NC 27599, O: 919.962.6977, [amorganlopez@unc.edu](mailto:amorganlopez@unc.edu).

<sup>6</sup>Posthumous

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

announcements emphasized the development and delivery of “community friendly” therapies including the use of open enrollment groups.

One of the more fundamental analytic challenges in modeling treatment outcomes from group treatment data is the issue of group membership turnover, which can be more problematic in open enrollment groups versus closed group (NIDA, 2003). For example, in closed enrollment groups, once the group is formed, no new members are added to the treatment group and any change in membership would only be due to termination, dropout or treatment completion. In this case, group membership is clear and can be modeled easily and defensibly in analytic frameworks that handle non-independence of observations emerging from individuals nested within treatment groups in a conventional manner under the hierarchical linear modeling framework. Yet closed enrollment groups are problematic in community settings, primarily because the groups need to wait for a requisite number of patients for the group to start; in the meantime, the treatment center loses billable hours and patients may opt for other alternatives if they deem the wait for treatment to be too long (Morgan-Lopez & Fals-Stewart, 2008a; Overholser, 2005; Yalom & Leszcz, 2005).

Conversely, in open enrollment treatment groups (OEGs), which recent data suggest are the norm in community settings (Fals-Stewart, 2005), members are continually and simultaneously added via new enrollments and removed via graduation, termination or dropout. Yet nearly all analyses in the generalized linear mixed modeling family, from repeated-measures ANOVA to latent growth models for non-normal outcomes, assume that the membership composition of the groups does not change during the life of the trial (Morgan-Lopez & Fals-Stewart, 2006). Because methods to model the impact of group membership turnover in treatment outcome analyses had not been developed at the time, and treatment researchers naturally wanted to avoid negative critiques from treatment review study sections, OEG trials were avoided.

Recently, there have been significant conceptual and empirical advances in quantitative methods for modeling data from community-friendly trials with OEGs, particularly from a missing data theory perspective (Morgan-Lopez & Fals-Stewart, 2006, 2007, 2008a, 2008b; Morgan-Lopez, Cluff & Fals-Stewart, 2009). In this work, the primary framework that has shown promise for modeling treatment outcome data from open enrollment groups in a defensible manner is latent class pattern mixture modeling (LCPMM; Lin et al., 2004; Roy, 2003). LCPMMs, primarily geared as a novel approach for modeling non-ignorable missingness provide a framework that more closely represents the process of turnover in group membership than traditional methods (e.g., latent growth models in structural equation modeling or equivalent repeated measures mixed models; MacCallum et al., 1997; Willett & Sayer, 1994) or even conventional pattern mixture models (Demirtas & Schafer, 2003; Hedeker & Gibbons, 1997).

LCPMMs handle the process of group membership turnover by allowing for fluctuations over time in the proportions of different types of latent attendance patterns and, therefore, session-to-session fluctuations in different subtypes of patients. The proportions of patients under each attendance class are allowed to vary at any given slice in time which the trial is running, which is consistent with changes over time/turnover in group composition

### **Statistical Power in Open Enrollment Group Trials**

While progress has been made in this relatively new area of methodological research, several concerns remain (Morgan-Lopez & Fals-Stewart, 2008a). One key concern is the estimation of statistical power in open enrollment group trial designs, be they newly-proposed trials or secondary analysis of existing OEG data. As the analytic and methodological barriers to the development of community-friendly open enrollment group

treatment trials have diminished, and as the number of grant applications involving OEG trials has increased (Onken, L.S., personal communication, 1 October 2008), guidance on the estimation of statistical power for OEG trials becomes paramount.

One of the more fundamental concerns in a power analysis, particularly as part of the critique of NIH grant applications, is the extent to which there is a (mis)match between the analytic model assumed in the power analysis and the actual analytic model used to analyze data once collected. Several decades ago, prior to the popularizing and availability of models under the generalized linear mixed modeling family (GLMM), the analytic options available to most treatment researchers for data analysis (e.g., t-tests, correlations,  $\chi^2$  contingency tables, analysis of variance) matched very well with the frameworks that were addressed in early texts on power (see e.g., Cohen, 1988) and readily available power analysis software such as SAS Proc POWER and SPSS SAMPLEPOWER. However, with the advent of methods under GLMM, power analyses in practice have not always mirrored the increase in complexity of the analyses planned, and ultimately executed.

Specifically, in our experience in reviewing NIH behavioral treatment trial applications, the most common discrepancy between the assumptions in the power analysis and in the proposed analysis plan is the failure to take into account non-independence of observations. The consequences for inference when failing to take non-independence into account in the analysis are fairly well-documented: increased Type I error rates when group-level variance components are not modeled when patients are nested within treatment groups (Baldwin, Murray & Shadish, 2005; Barcikowski, 1981; Hox, 2002) and increased Type II errors when the nesting of repeated measurements among individuals is not properly accounted for (Duncan, Duncan, Strycker & Li, 2002).

For contexts where the nesting hierarchy is unambiguous, approaches to power analysis based on analytic (Murray, 1998; Spybrook, Raudenbush, Congdon & Martinez, 2009) and Monte Carlo simulation methods (Muthén & Muthén, 2002) have been explicated. However, power analysis approaches for contexts where group membership is “fuzzy” or in constant flux are in need of development (Blalock, 1990); in fact, the development of new statistical methodologies and the subsequent lag in parallel development in power analysis approaches linked to those methodologies is not unusual (Duncan et al., 2002).

As the advocacy for the development of a community-friendly behavioral treatment portfolio remains salient (NIDA & NIAAA, 2003; NIDA, 2003a, 2003b; NIDA & NIH, 2008), the development of approaches to model data where continual group membership turnover is present in community-friendly trials should include parallel considerations for how to estimate statistical power for these community-friendly trials. The crux of the present article is to present a set of initial recommendations for estimating statistical power under a scenario where raw pilot data are not available and published summary data are the sole source from which parameters can be derived.

## Method

### Monte Carlo Simulation and Power Analysis

Since the design of open-enrollment treatment trials incorporates analytic and design complexities (e.g., group- and individual-level nesting, multiple attendance patterns, changes in group membership) that are neither accounted for in many closed form power analysis software packages nor power tables, the use of specially-tailored Monte Carlo simulations for the estimation of statistical power are an option that offers considerable flexibility (Muthén & Muthén, 2002).

Simulations can be conducted in general statistical packages such as SAS and SPSS or in many “model-specific” packages (e.g., Mplus, EQS, or LISREL in structural equation modeling) using the random number generation facilities of each program. For a given an underlying population model, a researcher can study many properties of the statistical estimator and the model used to analyze sample data, whether or not the analysis model is specified as the same as the data-generating model. A non-exhaustive set of examples of the properties commonly examined in simulations include parameter estimate bias and confidence interval coverage (e.g., Collins et al., 2001). It is also not uncommon for the statistical power of newly-developed estimators to be evaluated (MacKinnon et al., 2002, 2004) or power to detect treatment effects for a proposed study (Muthén & Muthén, 2002) to be examined via simulation.

### Powering a New Secondary Analytic Project based on Published Data

It is well-known that the four general variables in a power analysis include a) the probability of rejecting a true null (the alpha level), b) the effect size, c) the sample size and d) the probability of rejecting a true null hypothesis (power). Depending on which factors are fixed for a given research context, a researcher will be executing a power determination analysis, a sample size determination analysis or a minimum detectable effect size (MDES) analysis (Spybrook et al., 2009). In contexts where sample size is fixed, such as for secondary analytic projects, a hybrid of a power determination and MDES analysis may be warranted to determine what is the MDES that will lead to power of .80 for sample size N and/or what is the power to detect particular value(s) of the effect size. In contexts where the sample size is to be determined in the analysis and the desired power is fixed at .80, a hybrid of the sample size and MDES analysis may be warranted.

In the current example, we illustrate both the power determination approach and the sample size determination approach using the power analysis process used for one of the grants that support this current article (R01DA025198, Morgan-Lopez, A.A., PI), a secondary analysis project to re-analyze data from what is currently the largest trial examining *Seeking Safety* (NIDA Clinical Trials Network protocol 0015; Hien, D.A., PI). *Seeking Safety* is an empirically-supported behavioral treatment of comorbid PTSD and substance abuse disorders among women who have experienced trauma (Hien, Nunes, Levin & Fraser, 2000; Hien, Cohen, Litt, Miele, & Capstick, 2004; Najavits, 2007). The sample size in the dataset is fixed at N = 353. However, we wish to also illustrate the sample size determination approach, as though sample size was *not* fixed, to mimic the process for power analyses for researchers determining sample size for new primary collection trials.

We wished to illustrate a situation where no raw pilot data were available<sup>1</sup>; all that was available were a set of published articles in the area of the grant application we are proposing upon which we would base our effect size(s) and other available parameters. The necessary steps for executing this power analysis would be a) identification of the necessary components for the power analysis model, b) estimation of individual-level fixed and random effects from the summary data in the article(s) for use as population parameters for treatment effects/effect sizes, c) solve for plausible values of group-level variance components, d) determine a plausible structure for class-specific treatment attendance/missingness patterns (i.e., treatment completers, dropouts, etc.), e) determine class-specific effect sizes across each attendance patterns using the original effect size from Step 3 as a base and f) execution of the simulation with the selected parameters. Prior to showing the illustration, we present the steps in greater detail.

<sup>1</sup>Though raw data would, of course, be available from the third author.

## Power Analysis Steps

**Step 1: Identify necessary components**—The first step in conducting the power analysis, as with any power analysis, is to identify the types of parameters specific to a power analysis for open enrollment groups. The components for a power analysis for open enrollment groups are many and include a) sample size, b) number of treatment groups, c) amount of anticipated missing data, d) differences in slopes over time on the outcome across treatment conditions (effect size(s)), e) group- and individual-level variance components, f) class-specific distributions of the point(s) of treatment entry, g) the amount of deviation from the overall effect size within each attendance class and h) the size of within-individual error variability. Data for some components will be available from published articles and, as in many power analyses, plausible values would have to have reasonable justification for their use in the absence of available data for the particular component.

**Step 2: Estimate individual-level fixed and random effects under LGM from summary data**—Summary data from tables in published articles provide a reasonable source of information from which to estimate fixed effects and perhaps random effects as well (assuming that fixed and random effects are not already reported in the article as part of a growth modeling analysis; if so, they can be used as population parameters). If fixed and random effect estimates are not already available, the task is to convert the means and covariances (or correlations and standard deviations) into fixed- and random-effect parameters for a population group-stratified latent growth model<sup>2,3</sup> (LGM); though the target model is a latent class pattern mixture model, the LGM gives us a good base by which to estimate the parameters that will not vary across attendance patterns in the power analysis as well as the parameters that will. From this model, we can then make decisions about characteristics of the LCPMM model (i.e., different within-class treatment effects, different patterns of attendance/missingness) in later steps.

Oftentimes information necessary for the derivation of the random effect parameters (i.e., covariances/correlations between repeated measures among individuals within groups, group-level variance components) is not available in the article and in many cases are no longer reported in articles due to journal space limitations. In this case, a correlation structure for the repeated measures on the outcome variable must be assumed and chosen with reasonable justification. Subsequent means, correlations and standard deviations for the treatment and control conditions are then used as summary data input in a two-group LGM model using any structural equation modeling software which accepts summary data as input. Typical considerations that impact model fit in latent growth models (e.g., non-linearity in change over time) need to be kept in mind just as would be the case in conventional model estimation.

**Step 3: Solve for group-level variance components**—The next step would be to select reasonable expected ratios of the variability in the growth parameters that are due to group-level nesting, under the assumed study structure. In this case, the ratio of group-level variability to total variability in a parameter is:

$$\frac{\tau_{\beta}}{\tau_{\beta} + \tau_{\pi}} = \text{Intraclass Correlation} \quad (1)$$

<sup>2</sup>The parameters from the initial group-nested latent growth model would serve as the basis for then building a population LCPMM model, where treatment effects vary across attendance/missing data classes.

<sup>3</sup>For group-nested data with few groups, it is recommended that nesting be handled under the assumption that the groups are sampled exhaustively via stratification as opposed to assumed that the groups are sampled from a universe of groups (i.e., clustering; see Asparouhov, 2005).



Where  $\tau_{\beta}$  is the group-level variance component for a parameter (e.g., intercept, slope) and  $\tau_{\pi}$  is the corresponding individual-level variance component. With a value for the intraclass correlation selected ahead of time and a value for  $\tau_{\pi}$  emerging from Step 2, one can solve for  $\tau_{\beta}$  algebraically.

**Step 4: Determine Attendance Class Structure**—Next, we determine the attendance class structure, where variation in class membership for is characterized by a) variation in outcome trajectories and treatment effects, b) variation in the probabilities of missingness/treatment attendance and c) differences in the distributions of calendar time with regard to when each individual case began treatment (Morgan-Lopez & Fals-Stewart, 2007, 2008b). With regard to the structuring of the joint outcome/attendance classes in such a power analysis, one must consider a) the number of classes, b) the proportions of cases within each class and c) the probability of missingness/attendance at each “timepoint” within each class; the names of these classes are typically derived from the patterns of these probabilities over time (Morgan-Lopez & Fals-Stewart, 2007). In recent applications of LCPMM in multiple datasets, there have been three classes estimated with class membership jointly determined by attendance patterns, treatment outcomes and point of treatment entry (Hien, Morgan-Lopez, Campbell, Saavedra et al., under review; Morgan-Lopez & Fals-Stewart, 2007). In each case, there was a class of patients who had consistently high probabilities of attending treatment each session (Completers) and a class that, at a certain point, had virtually 0 probability of attending treatment (Dropouts); a third class emerged in each case with a set of probabilities over time for treatment attendance that were higher than the Dropouts but lower than the Completers so this three class structure will be used in the current illustration.

**Step 5: Determine class-specific effect sizes within each attendance pattern**—A notable characteristic observed in open enrollment data is the fact that differences in treatment effects have been observed that are dependent on the pattern of attendance. One finding that appears to be reasonably consistent is that the patients who tend to have optimal treatment outcomes in behavioral substance use treatment tend to self-titrate the dosage of their treatment to around half the available sessions (Barkham, Connell, Stiles, Miles, et al., 2006; Feaster, Newman & Rice, 2003; Hien et al., under review; Morgan-Lopez & Fals-Stewart, 2007, 2008b). As a result, in order to limit the scope of the power analysis and to be consistent with an emerging literature on treatment effects that are conditional on attendance, we recommend using a) the effect size found in Step 2 for the Completers class, b)  $\frac{1}{2}$  of the Completers effect size as the effect size for Dropouts and c) twice the effect size for Completers would serve as the effect size for a class with “erratic” attendance.

We must first calculate the effect sizes from the base parameters. Several potential options for effect size calculation in the context of latent growth models are available including variants of standardized mean differences in slopes (Feingold, 2009; Muthén & Curran, 1997; Raudenbush & Liu, 2001) and  $R^2$  measures of the variance accounted for in latent slope(s) as a function of treatment (Muthén & Muthén, 1998-2009). However, in power analyses for latent growth models/longitudinal HLMs, Raudenbush & Liu (2001) note that both the effect size *and* the precision of the measurement of the latent curves need to be considered jointly; the concern is that even if other parameters (e.g., effect size, sample size) were held constant in the power analysis, one can always increase statistical power by increasing the reliability of the curves (i.e., by decreasing the value of the within-individual residual variance). However, if this residual variance is set too low relative to the real value in the population (i.e., higher reliability), the study may ultimately be underpowered not because the effect size was over estimated but because the curve reliability may be lower in reality than what was assumed in the power analysis.

Calculation of the overall effect size in-treatment slope from the Hien et al., (2004) data in Table 2 uses the standardized effect size for group-nested longitudinal data (Raudenbush & Liu, 2001; Spybrook et al., 2009):

$$\frac{\beta_p}{\sqrt{\tau_{\pi 1} + \tau_{\beta 1}}} \quad (2)$$

where  $\beta_p$  is the unstandardized treatment slope of interest from Step 2,  $\tau_{\pi 1}$  is the individual-level variance of the corresponding slope from Step 2 and  $\tau_{\beta 1}$  is the group-level variance of the corresponding slope solved for in Step 3

The formula for the reliability of the growth curves (Raudenbush & Liu, 2001) is:

$$\frac{\tau_p}{\tau_p + V_p} \quad (3)$$

where  $\tau_p$  is the variance of the slope of interest and  $V_p$  is an expression of the total variability among all observations that is *not* attributable to “true” variability among individual slopes.  $V_p$  is expressed as<sup>3</sup> (Spybrook et al., 2009):

$$\frac{\sigma^2(f)(M-2)!}{1/12(M+1)!} \quad (4)$$

Where  $f$  is the number of observations per unit time (e.g., a “0, .5, 1...” time structure would have 2 observations per unit of time),  $\sigma^2$  is the within-individual residual variance across all timepoints and  $M$  = the number of timepoints + 1.

**Hypothetical Context for the Power Analysis**—The simulation mimics the design of the NIDA Clinical Trials Network six-site study (NIDA CTN Protocol 0015, Hien, D.A., PI) comparing *Seeking Safety* (SS) and a Women’s Health Education comparison condition. Both treatments were delivered in the group modality with rolling enrollment, with each of six sites running an SS group and a WHE group. Each of the rolling treatment groups could have 3 or more “members” on any given week across a period of 20 “months”.

**Model Overview**—In this Monte Carlo power analysis, the population model is a three-class, five timepoint (group-nested) LCPMM model as shown in Figure 1. The focal variable in the model is “Attend<sub>k</sub>”, a latent class variable capturing unobserved heterogeneity across three sets of variables: a) binary indicators of attending a treatment session ( $A_2$ - $A_4$ ) or attending a follow-up assessment ( $A_{1Wk}$ - $A_{12m}$ ), b) differences in changes over time in substance use ( $Y_{base}$ - $Y_{12m}$ ) as captured by SS/WHE differences in growth parameters ( $\beta_{ITSi}$ ,  $\beta_{PTSi}$ ,  $\alpha_{PTTi}$ ) which constitute treatment effects that are specific to each attendance class and c) variation in the distribution of the month in which each patient started treatment (StartMonth).

A piecewise linear structure is assumed for the outcome trajectories within each of the three classes such that there are two periods of growth of interest: a) changes in substance use from pre-treatment to 1-week post-treatment and b) changes from 1-week post-treatment through 1-year post-treatment; timesteps in the population model were set such that the intercept captured the estimated level of the outcome variable at the last time point (i.e., “1-year follow-up”). Treatment effect sizes were set to differ across the three classes (see section of Class-Specific Treatment Effects below).

## Illustration

**Step 1: Identify necessary components**—In the present case, we have published data from Hien et al., (2004) from which we can draw information relevant to some of the parameters. The key sets of information available from the Hien et al., (2004) article are the means and standard deviations over time on Substance Use Severity (Weiss, Hufford, Najavits, & Shaw, 1995) for women in the *Seeking Safety* and Community Care (comparison) conditions across four waves (baseline, post-treatment, 6- and 9-month follow-up; Table 2 of Hien et al., 2004, page 1429) and a conservative treatment completion rate of 60% (which will inform the attendance class membership probabilities; actual was 75%).

In the absence of information on the correlation structure, we assumed for the main power analyses a) a correlation structure among the repeated measures in the population that was consistent with our experiences in treatment outcome studies: correlations of .5 for adjacent timepoints (e.g., Time<sub>1</sub>-Time<sub>2</sub>, Time<sub>2</sub>-Time<sub>3</sub>), .3 for repeated measures which were two timepoints removed (e.g., Time<sub>1</sub>-Time<sub>3</sub>) and .1 for the correlation between Time<sub>1</sub> and Time<sub>4</sub>. The actual means and standard deviations from Hien et al., (2004) and the assumed correlations are shown in Table 1.

**Step 2: Estimate individual-level fixed and random effects under LGM from summary data**—The data in Table 1 were used as input data for the estimation of a 2-group LGM model. A piecewise-linear model was specified, with timesteps structured such that the two periods of growth were a) from baseline to post-treatment (in-treatment slope; ITS) and b) from post-treatment through 9-month follow-up (post-treatment slope; PTS); note also that Time = 0 was set to the 9 month follow-up assessment. In this case, the initial period of growth is modeled with only two timepoints which would lead to a problem of model underidentification unless additional constraints are made. The parameters in this model can only be estimated uniquely when the residual variances of the within-individual model are constrained to equality across time (and across groups).

Other random effect parameters (e.g., variances and covariances among the intercept, ITS and PTS) are also constrained to equality across groups. The growth parameter means (intercept, ITS and PTS) are allowed to vary across groups; however, in order to estimate the mean differences across groups in a single step, additional parameters need to be specified outside of the model in the multiple-group LGM framework (e.g., the Model Constraint option in Mplus 5.2). The program used for this step is included in Appendix A.

The values of these differences constitute the fixed effect regression parameters (linking treatment condition to differences in growth parameters) to be used subsequently in the Monte Carlo power analysis. This model, however, did not fit the data well (to the extent that the data are modeled with arbitrary correlations),  $X^2(15) = 20.323$ , CFI = .93, RMSEA = .08 (90% CI: .00-.16). The source of the model misfit was the assumption of the equality of the within-individual residual variances across treatment conditions. Once this constraint was relaxed across treatment conditions (but remained equated across-time *within* the two treatment conditions), the model fit the data well in terms of stand-alone model fit ( $X^2(14) = 10.207$ , CFI = 1.0, RMSEA = .00 (90% CI: .00-.09) and in comparison to the more restrictive model  $\Delta X^2(1) > 10$ ,  $p < .001$ . It would be ideal to have a single pooled within-individual residual variance; the model, based in part on data from Hien et al., (2004), suggests that this assumption may not hold, most likely due to the differences in standard deviations between Seeking Safety and Community Care (see Table 1). Parameter estimates from the LGM model, which would be used as population parameters in the subsequent simulation, are shown in Table 2.



**Step 3: Solve for group-level variance components**—In the Monte Carlo simulation a) there are six treatment groups per condition as is the case in the larger Seeking Safety trial and b) the level of group-level variability in growth parameters is non-zero. As the treatment arms in Hien et al., (2004) were all delivered in the individual modality, no estimates of the group-level variance components are available from the benchmark study. There is limited guidance from the literature on the size of group-level variance components relative to the total variability (group- and individual-level) among growth parameters in the context of open enrollment trials; it has been suggested that the ratio of group-to-total variability in growth parameters is lower than in closed group trials (Morgan-Lopez & Fals-Stewart, 2006) with estimates ranging between .01-.02 (Morgan-Lopez & Fals-Stewart, 2007). With this in mind, we set out to calculate the values of the group-level variance components that would yield a ratio of groups-to-total variability in growth parameters of .02.

Using Equation 1, we solve for the value ( $g$ ) that would yield a groups-to-total variability ratio of .02 given an individual-level variance for the in-treatment slope of .201 (See Table 2):

$$\frac{g}{g+.201}=.02$$

In this case,  $g = .0041$ ; a similar process were conducted to solve for group-level variability in the post-treatment slope and post-treatment intercept for the power analysis.

**Step 4: Determine Attendance Class Structure**—We use information from both Hien et al., (2004) and Morgan-Lopez and Fals-Stewart (2007) to guide the attendance class structure. First, Morgan-Lopez and Fals-Stewart (2007) found a 3-class solution was optimal, with an attendance pattern structure where one class had between a 74-93% probability of attending each treatment session (Completers), a class which had a monotonically decreasing probability of attending treatment over time which decreased to near 0 before the end of the treatment protocol (Dropouts) and a third class where the pattern of attendance varied wildly throughout treatment (Erratics). While the Completers class accounted for 60% of the sample in Morgan-Lopez and Fals-Stewart (2007), Hien et al., (2004) report that 75% of women completed treatment; we decided to structure a conservative rate for treatment completion in the proposed population structure with 3 classes (Completers, Dropouts, Erratics) with a 60%/20%/20% split respectively among the classes.

**Step 5: Determine class-specific effect sizes within each attendance pattern**—Recall that the standardized effect size for group-nested longitudinal data (Raudenbush & Liu, 2001; Spybrook et al., 2009):

$$\frac{\beta_p}{\sqrt{\tau_{\pi 1} + \tau_{\beta 1}}} \quad (2)$$

where  $\beta_p$  is the unstandardized treatment slope of interest from Step 2 (-.209 in Table 2),  $\tau_{\pi 1}$  is the individual-level variance of the corresponding slope from Step 2 (.201 in Table 2) and  $\tau_{\beta 1}$  is the group-level variance of the corresponding slope solved for in Step 3 (.0041 in Table 2); the standardized ES in this case is -.461, considered slightly below a medium slope effect showing reduced substance use for *Seeking Safety* compared to *Community Care* from pre-treatment to treatment termination (Cohen, 1988); this ES was used for the Completers class treatment effect. An effect size of -.230 ( $\beta_p = -.105$ ) was used for the

population value for the Dropout class treatment effect and an effect size of  $-.922$  ( $\beta_P = -.418$ ) was used for the Erratics class treatment effect.

The formula for the reliability of the growth curves (Raudenbush & Liu, 2001) is:

$$\frac{\tau_p}{\tau_p + V_p} \quad (3)$$

where  $\tau_p$  is the variance of the slope of interest and  $V_p$  is an expression of the total variability among all observations that is *not* attributable to “true” variability among individual slopes.  $V_p$  is expressed as<sup>3</sup> (Spybrook et al., 2009):

$$\frac{\sigma^2(f)(M-2)!}{1/12(M+1)!} \quad (4)$$

Where  $f$  is the number of observations per unit time (e.g., a “0, .5, 1...” time structure would have 2 observations per unit of time),  $\sigma^2$  is the within-individual residual variance across all timepoints and  $M$  = the number of timepoints + 1. Since  $f = 1$ ,  $M = 3$  and  $\sigma^2 = .204$  for the in-treatment slope,  $V = .102$  and the slope reliability =  $.663$ .

## Population Parameters

**Attendance/Assessment Patterns**—The classes are also distinguished by differences in the probabilities of attending treatment sessions during the “in-treatment” phase (not including the 1<sup>st</sup> observation) or post treatment assessments. These conditional probabilities are described as follows: (a) *Consistent Attenders*: set to have a 90% probability of “attendance/compliance” across the final six assessments; (b) *Dropouts*: set to have the following probabilities of attendance across the final six assessments: (.90, .70, .40, .40, .40, .20, .10) and (c) *Erratics*: set to have the following probabilities of attendance across: (.20, .20, .80, .80, .20, .20, .80). It is noted that if a 0 is generated based on the conditional probabilities for any given case in a simulated sample at a timepoint where there is a corresponding measure of the outcome variable (e.g., baseline, termination, post-treatment assessments), then the corresponding value on the outcome ( $Y_T$ ) was set to missing.

**Class-Specific Treatment Effects<sup>4</sup>**—Class-specific population treatment effects, defined as mean differences between *Seeking Safety* and WHE on two key growth parameters (i.e., intercept at 1-year follow-up, slope from baseline to 1-week follow-up). Three different effect sizes for the in-treatment slope differences were used within the three classes: a) the *actual* effect size for the in-treatment slope for the Completers class from the parameters in Table 2 ( $\beta_P = -.209$ , standardized mean difference =  $-.461$ ), b) *twice* the in-treatment effect for the Erratics class ( $\beta_P = -.418$ , standardized mean difference =  $-.922$ ) and c) *half* the in-treatment effect for the Dropouts class ( $\beta_P = -.105$ , standardized mean difference =  $-.230$ ). The post-treatment slopes and the post-treatment intercept were held constant across classes to the values shown in Table 2.

**Class Proportions**—The proportions of the population from each attendance class were fixed to 60% Completers, 20% Dropouts and 20% Erratics.

<sup>4</sup>Recent concerns about the variability in effect sizes from pilot studies have been raised (Kraemer, Mintz, Noda, Tinklenberg & Yesavage, 2006). Kraemer and colleagues recommend using effect sizes for power analyses that are the minimum effect sizes to be of clinical interest; in this case, we argue that the effect size(s) obtained from the parameters derived from Hien et al., (2004) are of clinical interest, given a primary effect size slightly below a treatment slope difference that would be considered medium.

**Sample Sizes**—Sample sizes of 150, 250, 353 (the actual sample size in the current Seeking Safety CTN trial) and 450 were used.

**Simulation Heuristics**—First, simulated data were generated in SAS v9 under a 3-class latent class pattern mixture population structure, with population values corresponding to the class-specific treatment effects and class proportions as listed above; 1000 replications were generated under each specific sample size. Once generated, each of the 1000 datasets was analyzed in Mplus v5.2 in the External Montecarlo analysis framework (Muthén & Muthén, 1998-2008) under maximum likelihood estimation for non-normal data and/or non-independent observations under stratification (Asparouhov, 2005). Each dataset was analyzed under a correctly-specified 3-class LCPMM model; power was also explored under conventional growth modeling for comparison.

The weighted averaged treatment effect estimates under LCPMM were calculated by first converting the multinomial logit parameter estimates in the model to estimated proportions of class membership. Next, the proportions were used as weights against which the weighted-averaged treatment effect and delta method standard errors were estimated using the Model Constraint command in Mplus (see Hedeker & Gibbons, 1997, p.74-76). Across each simulated sample, the proportion of times out of the total number of replications that the weighted averaged treatment effect was significantly different from 0 was the observed power to detect the effect.

## Results

The results for statistical power for class-specific and weighted-averaged treatment effects are shown in Table 3. Power estimates were examined in concert with other measures for assessing the quality of model performance such as confidence interval coverage and standardized bias (Collins, Schafer & Kam, 2001). Coverage is defined as the proportion of times the population parameter falls within the sample confidence intervals across each replications; the ideal value is .95 and values at or below .90 are problematic. Standardized bias is defined as the difference between the population parameter and the average estimate across all replications divided by the standard deviation of the estimates. Collins et al., (2001) recommend values exceeding  $|\cdot40|$  as problematic.

Recall that the basic parameters for the power analysis were a) treatment effect sizes of  $-.46$  for Completers,  $-.92$  for Erratics and  $-.23$  for Dropouts, b) ICC(s) of  $.02$  and c) Growth Curve Reliabilities of  $.66$  for the in-treatment slope. Under LCPMM, power would only approach  $.80$  for overall/weighted-averaged treatment effects until the sample size was raised to 450; for the *actual* sample size of  $N = 353$  for the Seeking Safety dataset, power was  $.678$ . Coverage and standardized bias were at acceptable levels.

Power was also examined under the LGM framework with the same simulated datasets. Power under LGM reached  $.799$  for an  $N$  of 450 and  $.737$  for the actual  $N$  of 353. Despite estimating power under a misspecified LGM model, standardized bias and coverage rates were acceptable under LGM. However, it is noted that one of the conditions where LGMs will not yield biased estimates and perform similarly to LCPMMs when data were generated under LCPMM are when the class with intermediate missingness a) is the smallest class and b) has the largest effect size; in a majority of other combinations of class structure, class size and effect size, LGMs and (LC)PMMs can yield different inferences regarding overall treatment effects which has been shown in simulated (Demirtas & Schafer, 2003) and real data (Morgan-Lopez & Fals-Stewart, 2007).

To illustrate this, we simulated data with the exact same properties except a) treatment effect coefficients corresponding to the class-specific effect sizes from Morgan-Lopez and Fals-Stewart (2007) were used and b) the class proportions were altered such that the patterns were: 55% Completers, 35% Dropouts and 10% Erratics. First, the class-specific effect sizes from Morgan-Lopez and Fals-Stewart (2007; Cohen's  $d$ s for Completers = 0, Dropouts = .85 and Erratics = -1.5) were converted to path coefficients for the difference in in-treatment growth for a given level of growth parameter variance using Equation 2 and solving for  $\beta$ .  $\beta$ s, given the specified group- and individual-level variances were 0 for Completers, .3849 for Dropouts and -.6793 for Erratics. Weighting the parameters by the class proportions yields an overall  $\beta$  of .0667. Data were simulated and analyzed under sample sizes of 150, 250, 353 and 450 (1000 replications each) just as for the original analysis.

As shown in Table 4, under a properly-specified 3-class LCPMM, standardized bias rates hovered around the bounds of what is considered un/acceptable. Confidence interval coverage did not decrease below .90. Under the LGM framework, standardized bias rates were greater than twice the rate considered problematic as sample size exceeded 350. Also under LGM, coverage rates generally were lower than .90.

## Discussion

Within the last five to ten years, federal agencies and community treatment providers have called for greater ecological validity in the designs of proposed behavioral treatment trials. While several factors have accounted for the disconnect between treatment research and treatment-in-practice (Barkham & Mellor-Clark, 2003; Greene, 2004; NIDA, 2003), analytic difficulties in modeling data from designs resembling treatment-in-practice (i.e., open enrollment groups) had been highlighted as a major concern (Weiss et al., 2004). As the development of methodologies to handle data from open enrollment groups has progressed (Bauer, Gottfredson & Morgan-Lopez, 2009; Morgan-Lopez & Fals-Stewart, 2007, 2008b), and defensible analytic options have become available, the number of open enrollment group trials proposed in behavioral treatment grant applications has increased. However, to our knowledge, all of these proposed trials had power analyses that were incompatible with the planned analysis, a phenomenon that is common when new analytic frameworks are introduced (Duncan et al., 2002).

In this article, we illustrated an exemplar for the estimation of statistical power for the recently-developed latent class pattern mixture model (Lin et al., 2004; Roy, 2003) which may be parameterized to capture and model the impact of membership turnover in treatment groups. The match between the analytic framework described in an analysis section of an NIH grant and the analytic framework that underlies the power analysis for the study is critical; if not present the justification of any trial can be compromised, as the study may be powered based on faulty assumptions.

The interest in this power analysis was primarily to estimate statistical power for overall treatment effects when treatment effects, particularly on the in-treatment slope, vary across attendance classes. In powering a trial of the size of the *Seeking Safety* NIDA CTN protocol, it is much more typical (and ideal) to have raw pilot data available from which to base *all* population parameters for a power analysis, though cautions have been levied against balancing effect sizes from pilots against powering studies to effect sizes that are clinically meaningful (Kraemer et al., 2006). Historically this has been in the context of the Stage Model for Behavioral Therapies Research (Carroll & Nuro, 2002; Rounsaville, Carroll & Onken, 2001) where data from Stage I a/1b trials are used explicitly to test feasibility for Stage II/III trials. Furthermore, this power analysis was geared to mimic the process of sample size estimation for a newly-planned trial and illustrate power estimation for an

already-fixed sample size for *secondary analyses* of existing data; however, as there has been an increase in interest in secondary analysis of treatment and health services datasets (e.g., PA-07-113; NIDA, 2007) this situation will not be uncommon.

Several aspects of this power analysis required a bit of educated guesswork that would not otherwise be necessary with raw pilot data. For fixed-effect parameters, we could rely on modeling the mean differences over time between the SS condition and the Community Care comparison condition from Hien et al., (2004). Yet we had no knowledge of the class structure (i.e., number of classes, class proportions, attendance patterns) in the data from the larger *Seeking Safety* trial. We also did not have any guidance on the random effect structure; we had a limited understanding of the size of group-level variance components in open enrollment data from the behavioral treatment literature and little-to-no specific guidance on the correlation structure among repeated measures (which contribute directly to the *individual*-level random effects). As a result, we worked with a small range of plausible parameters, each of which could have a small (e.g., number of individuals within cluster/strata) or large impact (e.g., group ICC, effect size) on power (Spybrook et al., 2009).

In this power analysis demonstration, we focused primarily on treatment effects on the in-treatment slope (as the parameter estimate represented a medium effect). The results of the power analysis suggested that, for treatment effects on the in-treatment slope, power to detect the overall in-treatment treatment effect would approach .80 when sample size approaches 450. This would seem to present a quandary of sorts, because the effect sizes correspond both to effect sizes from previous research on *Seeking Safety* and effect sizes that are clinically meaningful according to Kraemer et al., (2006) but require a sample size that exceeds the upper end of the range of sample sizes typically observed in NIDA- and NIAAA-funded behavioral treatment trials; in our experience as reviewers of behavioral treatment trial applications, these trials rarely are funded at Ns above 350.

An additional consideration in structuring such a power analysis may be the basics on the type of comparison condition. The original power analysis examined in this paper was based on a comparison between an active treatment (*Seeking Safety*) and an attention control condition (Hien et al., 2004). However, this is contrasted against Morgan-Lopez and Fals-Stewart (2007) where two active treatments were compared in different modalities (individual versus group), which may be more likely to lead to conflicting inferences across attendance classes – and thus different inferences across the types of analyses. This leads to an interesting irony: the conditions that may be most likely to lead to different inferences between LCPMM and LGM may be the least conducive to statistical power. Additional studies using the LCPMM framework will shed light on which class structure may be most likely to be observed in treatment outcome studies regarding ordering and sizes of effect sizes across classes: one that is maximally likely to lead to different results or a structure that may yield similar results between LCPMMs and LGMs.

## Conclusion

Despite the call from federal agencies and community treatment providers for ecological validity in drug and alcoholism treatment research, analytic challenges of rolling group/open enrollment data has hindered research in this area (NIDA, 2003; NIDA & NIH, 2008). As analytic and methodological tools become available to behavioral treatment researchers, it has increased the volume of NIH grant submissions incorporating ecologically-consistent designs which will hopefully lead to a bridging of the gap between the treatment research portfolios at NIDA, NIAAA and other relevant agencies. But as the number of applications for these types of trials increases (particularly those who proposed the use of latent class pattern mixture models), there have been discrepancies between the planned analysis model and the model assumed in the power analysis; we assume that this is because the analysis

framework is relatively novel and power analysis examples for new frameworks typically lag behind illustrations of the frameworks themselves (Duncan et al., 2002). Hopefully this step-by-step power analysis demonstration for data from open enrollment groups will provide researchers within and outside of the behavioral treatments area with guideposts for powering new trials and secondary analytic projects which will further bridge the gap between treatment research and treatment-in-practice.

## Acknowledgments

This project was supported by grants R01DA025198, R21DA021147 and R21AA016543 (Antonio A. Morgan-Lopez, PI), grants R01DA12189, R01DA014402, R01DA014402-SUPL, R01DA015937, R01DA016236, R01DA016236-SUPL and the Alpha Foundation (William S. Fals-Stewart, PI), grants R01AA014341, R01DA23187 and NIDA CTN Protocol 0015 (Denise A. Hien, PI), and a supplement to grant R01DA025198 and a Professional Development Award from RTI International (Lissette M. Saavedra, PI).

## Appendix A

### Converting Means, Correlations and Standard Deviations to Model Parameters (using data in Table 1)

```

DATA:
FILE IS "C:\Users\aaaml\My Documents (RTI)\Q Drive Backup\001 NIDA - RGG I
\Paper 7 - Power\dahmcovA2.dat";
TYPE IS MEANS fullcov;
NGROUPS = 2;
NOBSERVATIONS = 53 54;
VARIABLE:
NAMES ARE x1 x2 x3 x4;
USEVARIABLES ARE x1-x4;
ANALYSIS:
TYPE IS meanstructure;
ESTIMATOR IS ML;
ITERATIONS = 1000;
CONVERGENCE = 0.00005;
MODEL:
a b1 | x1@-1 x2@0 x3@0 x4@0;
a b2 | x1@-2 x2@-2 x3@-1 x4@0;
Model g1:
[a](a);
[b1](b);
[b2](c);
a(1);
b1(2);
b2(3);
a with b1(4);
a with b2(5);
b1 with b2(6);
x1-x4(7);
Model g2:
[a](d);
[b1](e);

```



```

[b2](f);
a(1);
b1(2);
b2(3);
a with b1(4);
a with b2(5);
b1 with b2(6);
x1-x4(8);
Model constraint:
NEW(intdif sldif s2dif);
intdif = a-d;
sldif=b-e;
s2dif=c-f;
OUTPUT: SAMPSTAT standardized;

```

## Appendix B

### SAS Program for Generating Group-Stratified LCPMM Data

```

%let d=.dat;
%macro
q(set,combo,egit,r0gi,r1gi,r2gi,a000,b100,b200,g001,g101,g201,p0g,plg,p2g
,a00c,b10c,b20c,g01c,g11c,g21c,a00d,b10d,b20d,g01d,g11d,g21d,
a00e,b10e,b20e,g01e,g11e,g21e); *Macro variables for all parameters;
%do i=1 %to &set;
data a;
do f=1 to 12; *12 Treatment Groups
group=f;
x=ranbin(2,1,.5); *Seed of 2 yields a draw of 6 treatment and 6 control
groups across each replication;
*Group Randomization to Conditions with equal probability;
int=&a000+&g001*x+rannor(0)*(sqrt(&p0g));
*Group-level random intercept. Treatment Effect = -.261, Variance of .0045
(see end of macro/Table 2);
slopel=&b100+&g101*x+rannor(0)*(sqrt(&plg));
*Group-level random in-treatment slope. Treatment Effect = -.209. Variance
of .0041;
slope2=&b200+&g201*x+rannor(0)*(sqrt(&p2g));
*Group-level random post-treatment slope. Treatment Effect = .110, Variance
of .0012;
output;
end;
data b;
do g=1 to 353;
*Generate 353 individuals within 12 treatment groups;
id=g;
group=1+round((12-1)*ranuni(0),1);
output;
end;

```

```

run;
proc sort data=b;
by group id; run;quit;
proc sort data=a;
by group;run;quit;
data c; merge a b; by group;
/*Multinomial logits which yield a 60/20/20 split*/
z1=1.101;
z2=.00323;
p1 = exp(z1)/(1+ exp(z1)+ exp(z2));
p2 = exp(z2)/(1+ exp(z1)+ exp(z2));
p3 = 1/(1+ exp(z1)+ exp(z2));
/*Class membership draws with 60/20/20 probability*/
class = rantbl(0, p1, p2);
/*Individual Random Effects - correspond to .222, .201, .058 (see Table 2)*/
u0=rannor(0)*sqrt(&r0gi);
u1=rannor(0)*sqrt(&r1gi);
u2=rannor(0)*sqrt(&r2gi);
/*Completers - always @ 60% of the mixture*/
if class=1 then alpha=&a00c+&g01c*x+u0;
if class=1 then betal=&b10c+&g11c*x+u1;
*Class-specific deviation in the treatment effect: 0 for completers;
if class=1 then beta2=&b20c+&g21c*x+u2;
if class=1 then m2=ranbin(0,1,.9);
*Probabilities of showing up for treatment/assessment = 90% for each
observation;
if class=1 then m2a=ranbin(0,1,.9);
if class=1 then m2b=ranbin(0,1,.9);
if class=1 then m3=ranbin(0,1,.9);
if class=1 then m4=ranbin(0,1,.9);
if class=1 then m5=ranbin(0,1,.9);
if class=1 then m6=ranbin(0,1,.9);
if class=1 then
startwk=rantbl(0,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,
.05,.05,.05,.05,.05);;
*Across months 1-20, equal probability of starting treatment in Months 1-20
(5%);
/*Droppers - always 20% of the mixture*/
if class=2 then alpha=&a00d+&g01d*x+u0;
if class=2 then betal=&b10d+&g11d*x+u1;
*Difference in treatment effect from overall treatment effect for Droppers
= .104 (-.209 versus -.105);
if class=2 then beta2=&b20d+&g21d*x+u2;
if class=2 then m2=ranbin(0,1,.9);
*Probabilities of showing up for treatment/assessment = drop from 90% to 10%
over time;
if class=2 then m2a=ranbin(0,1,.7);
if class=2 then m2b=ranbin(0,1,.4);
if class=2 then m3=ranbin(0,1,.4);
if class=2 then m4=ranbin(0,1,.4);

```

```

if class=2 then m5=ranbin(0,1,.2);
if class=2 then m6=ranbin(0,1,.1);
if class=2 then
startwk=rantbl(0,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,.05,
.05,.05,.05,.05,.05);
*Across months 1-20, equal probability of starting treatment in Months 1-20
(5%);
/*Erratics - always 20% of the mixture*/
if class=3 then alpha=&a00e+&g01e*x+u0;
if class=3 then betal=&b10e+&g11e*x+u1;
*Difference in treatment effect from overall treatment effect for Erratics =
-.209 (-.418 versus -.209);
if class=3 then beta2=&b20e+&g21e*x+u2;
if class=3 then m2=ranbin(0,1,.2);
if class=3 then m2a=ranbin(0,1,.2);
if class=3 then m2b=ranbin(0,1,.8);
if class=3 then m3=ranbin(0,1,.8);
if class=3 then m4=ranbin(0,1,.2);
if class=3 then m5=ranbin(0,1,.2);
if class=3 then m6=ranbin(0,1,.8);
if class=3 then startwk=rantbl(0,.20,.0334,.03333333,.03333333,
.03333333,.03333333,.03333333,.03333333,.03333333,.03333333,
.03333333,.03333333,.20,.03333333,.03333333,.03333333,.03333333,
.03333333,.03333333,.03333333); /*Months "1" and "13" have a total of 40% of
the members of the Erratic Class*/
/*Generate Repeated Measures based on group- and individual-level fixed and
random effects*/
y1=alpha+int+(slope1+beta1)*-1+(slope2+beta2)*-1+rannor(0)*(sqrt(&egit));
y3=alpha+int+(slope1+beta1)*0+(slope2+beta2)*-1+rannor(0)*(sqrt(&egit));
y4=alpha+int+(slope1+beta1)*0+(slope2+beta2)*-.67+rannor(0)*(sqrt(&egit));
y5=alpha+int+(slope1+beta1)*0+(slope2+beta2)*-.33+rannor(0)*(sqrt(&egit));
y6=alpha+int+(slope1+beta1)*0+(slope2+beta2)*0+rannor(0)*(sqrt(&egit));
/*Impose missingness*/
if m3=0 then y3=999;
if m4=0 then y4=999;
if m5=0 then y5=999;
if m6=0 then y6=999;
data c; set c;
array cat y1--y6;
do over cat;
cat=round((cat),.001);
end;
run;
data c; set c;
*if _imputation_=&imp;
file "C:\Users\aaml\My Documents (RTI)\Q Drive Backup\001 NIDA - RGG I\Paper
7 - Power\rnrcombo&combo\test&i&d";
/*Save ASCII datasets for external monte carlo analysis in Mplus*/
put @1 group @6 x @20 m2 @25 m2a @30 m2b @35 m3 @40 m4 @45 m5 @50 m6 @55
startwk @60 y1 @100 y3 @110 y4 @120 y5 @130 y6;

```

```

run;
%end;
%mend;
%q(set=1000,combo=1,egit=.204,
r0gi=.222,r1gi=.201,r2gi=.058,
a000=0,b100=0,b200=0,
g001=-.261,g101=-.209,g201=.110,
p0g=.0045,p1g=.0041,p2g=.0012,
a00c=0,b10c=0,b20c=0,
g01c=0,g11c=0,g21c=0,
a00d=0,b10d=0,b20d=0,
g01d=0,g11d=.104,g21d=0,
a00e=0,b10e=0,b20e=0,
g01e=0,g11e=-.209,g21e=0);

```

## Appendix C

### Mplus program for External Monte Carlo Power Analysis

```

DATA:
FILE IS C:\Users\aaaml\My Documents (RTI)\Q Drive Backup\001 NIDA - RGG I
\Paper 7 - Power\rnrcomb01\test.dat;
type=montecarlo;
VARIABLE:
NAMES ARE group x m2 m2a m2b m3 m4 m5 m6 startwk
y1 y3 y4 y5 y6;
USEVARIABLES ARE x m2 m2a m2b m3 m4 m5 m6 startwk
y1 y3 y4 y5 y6;
stratification IS group;
!BETWEEN ARE x;
missing are all(999);
classes=miss(3);
categorical are m2 m2a m2b m3 m4 m5 m6;
ANALYSIS:
TYPE IS mixture missing complex;
! LOGHIGH = +15;
! LOGLOW = -15;
! UCELLSIZE = 0.01;
ESTIMATOR IS mlr;
H1ITERATIONS = 1000;
H1CONVERGENCE = 0.0001;
COVERAGE = 0.10;
LOGHIGH = +15;
LOGLOW = -15;
UCELLSIZE = 0.01;
LOGCRITERION = 0.0000001;
ITERATIONS = 1000;
CONVERGENCE = 0.000001;
MITERATIONS = 500;

```

```

MCONVERGENCE = 0.000001;
MIXC = ITERATIONS;
MCITERATIONS = 2;
MIXU = ITERATIONS;
MUITERATIONS = 2;
starts = 0;
processors=4;
information=observed;
link=probit;
MODEL:
%overall%
ba bb1 | y1@-1 y3@0 y4@0 y5@0 y6@0;
ba bb2 | y1@-1 y3@-1 y4@-.67 y5@-.33 y6@0;
ba on x;
bb1 on x;
bb2 on x;
[miss#1*1.101] (logit1);
[miss#2*.00323] (logit2);
y1-y6*.204;
ba*.222;
bb1*.201;
bb2*.058;
%miss#1%
[ba*0 bb1*0 bb2*0];
[startwk*10];
startwk*25;
ba on x*-.261(a1);
bb1 on x*-.209(b11);
bb2 on x*.110(b21);
[m2$1-m6$1*-1.28];
%miss#2%
[ba*0 bb1*0 bb2*0];
[startwk*10];
startwk*25;
ba on x*-.261(a2);
bb1 on x*-.105(b12);
bb2 on x*.110(b22);
[m2$1*-1.28 m2a$1*-.524 m2b$1*.253 m3$1*.253 m4$1*.253 m5$1*.84 m6$1*1.28];
%miss#3%
[ba*0 bb1*0 bb2*0];
[startwk*9];
startwk*36;
ba on x*-.261(a3);
bb1 on x*-.418(b13);
bb2 on x*.110(b23);
[m2$1*.84 m2a$1*.84 m2b$1*-.84 m3$1*-.84 m4$1*.84 m5$1*.84 m6$1*-.84];
MODEL CONSTRAINT:
NEW(p1 p2 p3 awa*-.261 blwa*-.214 b2wa*.110 cvd cve dve);
p1 = exp(logit1)/(1+ exp(logit1)+ exp(logit2));
p2 = exp(logit2)/(1+ exp(logit1)+ exp(logit2));

```

```

p3 = 1/(1+ exp(logit1)+ exp(logit2));
awa = p1*a1 + p2*a2 + p3*a3;
blwa = p1*b11 + p2*b12 + p3*b13;
b2wa = p1*b21 + p2*b22 + p3*b23;
cvd = b11-b12;
cve = b11-b13;
dve= b12-b13;
OUTPUT: sampstat tech1 tech8 tech9;

```

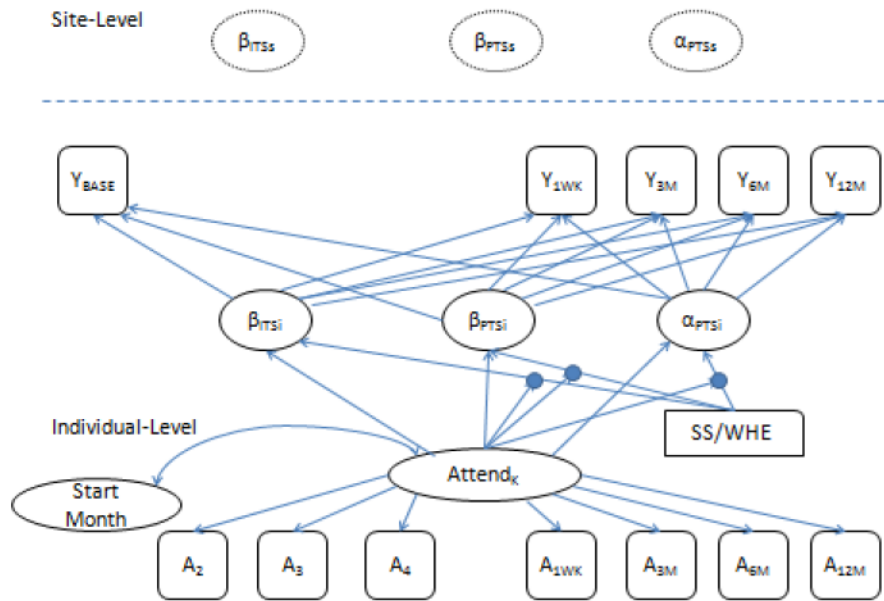
## References

- Asparouhov T. Sample weights in latent variable modeling. *Structural Equation Modeling*. 2005; 12:411–434.
- Baldwin SA, Murray DM, Shadish WR. Empirically supported treatments or type-1 errors?: A reevaluation of group-administered treatments on the empirically supported treatments list. *Journal of Consulting and Clinical Psychology*. 2005; 73:924–935. [PubMed: 16287392]
- Barcikowski RS. Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*. 1981; 6:267–285.
- Barkham M, Mellor-Clark J. Bridging evidence-based practice and practice-based evidence: developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy*. 2003; 10:319–327.
- Barkham M, Connell J, Stiles WB, Miles JNV, Margison F, Evans C, Mellor-Clark J. Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology*. 2006; 74:160–167. [PubMed: 16551153]
- Bauer DJ. A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*. 2005; 4:513–535.
- Bauer, DJ.; Gottfredson, NC.; Morgan-Lopez, AA. Toeplitz covariance structure for modeling group membership turnover in longitudinal studies. Presented at the UNC Quantitative Psychology Forum; Chapel Hill, NC. 20 February 2009; 2009.
- Blalock, HM. Auxillary measurement theories revisited. In: Hox, JJ.; De Jong-Gierveld, J., editors. *Operationalization and research strategy*. Amsterdam: 1990.
- Bryk, AS.; Raudenbush, SW. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, California: Sage Publications; 1992.
- Carroll KM, Nuro K. One size can't fit all: A stage model for psychotherapy manual development. *Clinical Psychology: Science & Practice*. 2002; 9:396–406.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.
- Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*. 2001; 6:330–351. [PubMed: 11778676]
- Demirtas H, Schafer JL. On the performance of random coefficient pattern mixture models for non-ignorable drop-out. *Statistics in Medicine*. 2003; 22:2553–2575. [PubMed: 12898544]
- Duncan TE, Duncan SC, Strycker LA, Li F. A latent variable framework for power estimation within intervention contexts. *Journal of Psychopathology and Behavioral Assessment*. 2002; 24:1–12.
- Fals-Stewart, W. *Group-based couples therapy for drug abuse (R01DA016326)*. Bethesda, MD: National Institute on Drug Abuse; 2005.
- Feaster D, Newman F, Rice C. Longitudinal analysis when the experimenter does not determine when treatment ends: what is dose–response? *Clinical Psychology & Psychotherapy*. 2003; 10(6):352–360. [PubMed: 16724159]
- Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*. 2009; 14:43–53. [PubMed: 19271847]
- Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*. 1997; 2:64–78.



- Hien DA, Cohen LR, Litt L, Miele GM, Capstick C. Promising empirically supported treatments for substance-using women with PTSD: A randomized clinical trial comparing Seeking-Safety with Relapse Prevention. *American Journal of Psychiatry*. 2004; 161:1426–1432. [PubMed: 15285969]
- Hien DA, Nunes EV, Levin FR, Fraser D. PTSD and short-term outcome in early methadone treatment. *Journal of Substance Abuse Treatment*. 2000; 19:31–37. [PubMed: 10867298]
- Hien DA, Morgan-Lopez AA, Campbell ANC, Saavedra LM, Wu E, Cohen LR. Can less be more?: A re-analysis of the CTN Women and Trauma trial. under review.
- Hox, J. *Multilevel analysis Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002.
- Lin HQ, McCulloch CE, Rosenheck RA. Latent pattern mixture model for informative intermittent missing data in longitudinal studies. *Biometrics*. 2004; 60:295–305. [PubMed: 15180654]
- MacKinnon DP, Lockwood CF, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*. 2002; 7(1):83–104. [PubMed: 11928892]
- MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*. 2004; 39(1):99–128. [PubMed: 20157642]
- MacKinnon DP, Taborga MP, Morgan-Lopez AA. Mediation designs for tobacco prevention research. *Drug & Alcohol Dependence*. 2002; 68:S69–S83. [PubMed: 12324176]
- Morgan-Lopez AA, Cluff LA, Fals-Stewart W. Capturing the impact of membership turnover in small groups via latent class growth analysis: Modeling the rise of the New York Knicks of the 1960's and 1970's. *Group Dynamics: Theory, Research and Practice*. 2009; 13:120–132.
- Morgan-Lopez AA, MacKinnon DP. Demonstration and evaluation of a method to assess mediated moderation. *Behavior Research Methods*. 2006; 38:77–87. [PubMed: 16817516]
- Morgan-Lopez AA, Fals-Stewart W. Analyzing data from open enrollment groups: Current considerations and future directions. *Journal of Substance Abuse Treatment*. 2008a; 35:36–40. [PubMed: 17936551]
- Morgan-Lopez AA, Fals-Stewart W. Consequences of misspecifying the number of latent treatment attendance classes in modeling group membership turnover within ecologically-valid behavioral treatment trials. *Journal of Substance Abuse Treatment*. 2008b; 35:396–409. [PubMed: 18513917]
- Morgan-Lopez AA, Fals-Stewart W. Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. *Journal of Consulting and Clinical Psychology*. 2007; 75:580–593. [PubMed: 17663612]
- Morgan-Lopez AA, Fals-Stewart W. Analytic complexities associated with group therapy in substance abuse treatment research: Problems, recommendations, and future directions. *Experimental & Clinical Psychopharmacology*. 2006; 14:265–273. [PubMed: 16756430]
- Morgan-Lopez AA, Hien DA, Fals-Stewart W, Wu E, Campbell AN. Latent class pattern mixture models and the recovery management paradigm. *Alcoholism: Clinical & Experimental Research*. 2009; 33:274A–274A.
- Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*. 1997; 2:371–402.
- Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*. 2002; 4:599–620.
- Muthén, LK.; Muthén, BO. *Mplus users guide*. 5. Los Angeles: Muthén & Muthén; 1998–2008.
- Murray, DM. *Design and analysis of group-randomized trials*. New York: Oxford; 1998.
- Najavits, LM. Seeking Safety: An evidence-based model for substance abuse and trauma/PTSD. In: Witkiewitz, KA.; Marlatt, GA., editors. *Therapists' Guide to Evidence-Based Relapse Prevention: Practical Resources for the Mental Health Professional*. San Diego: Elsevier; 2007. p. 141-167.
- National Institute on Drug Abuse (2003, April) Group therapy research. Workshop sponsored by the NIDA Behavioral Treatment Branch. Meeting summary available via World Wide Web at <http://www.drugabuse.gov/whatsnew/meetings/grouptherapy.html>

- National Institute on Drug Abuse, National Institute on Alcohol Abuse and Alcoholism. Request for applications for Group Therapy for Individuals in Drug Abuse and Alcoholism Treatment (RFA-DA-04-008). Washington, DC: Department of Health and Human Services; 2003a.
- National Institute on Drug Abuse, National Institute on Alcohol Abuse and Alcoholism. Behavioral Therapies Development Program (PA-03-126). Washington, DC: Department of Health and Human Services; 2003b.
- National Institute on Drug Abuse, National Institutes of Health. 2008Drugs, Brain, and Behavior: The Science of Addiction. NIH Pub No. 07-5605 Reprinted February 2008 Rockville MD. Available from [www.drugabuse.gov](http://www.drugabuse.gov)
- Overholser JC. Group Psychotherapy and Existential Concerns: An Interview with Irvin Yalom. *Journal of Contemporary Psychotherapy*. 2005; 35:185–197.
- Raudenbush SW, Liu XF. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*. 2001; 6:387–401. [PubMed: 11778679]
- Roy J. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*. 2003; 59:829–836. [PubMed: 14969461]
- Spybrook J, Raudenbush SW, Congdon R, Martinez A. 2009Optimal Design For Longitudinal and Multilevel Research: Documentation for the Optimal Design Software V.2.0. Available at [www.wtgrantfoundation.org](http://www.wtgrantfoundation.org)
- Weiss RD, Jaffe WB, de Menil VP, Cogley CB. Group therapy for substance use disorders: What do we know? *Harvard Review of Psychiatry*. 2004; 12:339–350. [PubMed: 15764469]
- Weiss, RD.; Hufford, C.; Najavits, LM.; Shaw, SR. Weekly Substance Use Inventory. Unpublished measure, Harvard University Medical School; Boston, MA: 1995.
- Willett J, Sayer A. Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*. 1994; 116:363–381. Yalom, I., editor. *The theory and practice of group psychotherapy*. 4. New York: Basic Books; 1995.
- Yalom, ID.; Leszcz, M. *The theory and practice of group psychotherapy*. Basic Books; 2005.
- Yuan, KH.; Bentler, PM. *Sociological Methodology* 2000. Washington, DC: American Sociological Association; 2000. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data; p. 165-200.



**Figure 1.** Latent Class Pattern Mixture Model.  $Attend_k$  = Latent Attendance Class Variable.  $Y_{BASE}$ - $Y_{12M}$  = Observed simulated outcome variable (e.g., past week substance use) from each individual’s baseline assessment through 1 year follow-up (noting that calendar time may be different for each “individual” for when they “came in for baseline”).  $A_2$ - $A_{12M}$  = Binary indicators of treatment attendance/assessment from the session after baseline through 1-year follow-up.  $SS/WHE$  = Treatment condition (Seeking Safety = 1; Women’s Health Education = 0).  $\alpha_{PTI_i}$  = estimated level of the outcome at time = 0 (i.e., 1-Year Follow-up).  $\beta_{ITS_i}$  = estimated rate of change from baseline to immediate post-treatment.  $\beta_{PTS_i}$  = estimated rate of change from immediate post-test through 1-year follow-up. Paths from “Attend” to the growth parameters indicate that the conditional means of the growth parameters vary across attendance class. Paths from “Attend” to the  $SS/WHE \rightarrow$  growth parameter links (as connected by the “dots”) indicate that the treatment effects vary across attendance classes.

**Table 1**

Input Means, Correlations and Standard Deviations from Hien et al., (2004)

	Seeking Safety		Community Care (comparison)					
	Baseline	Post-Tx	6m FU	9m FU	Baseline	Post-Tx	6m FU	9m FU
Correlations								
Base.	1				1			
Post-Tx	.5	1			.5	1		
6m FU	.3	.5	1		.3	.5	1	
9m FU	.1	.3	.5	1	.1	.3	.5	1
Means	-.08	-.15	-.12	-.08	.19	.36	.19	.21
SD	.68	.65	.61	.54	1.0	.78	.72	.76

Notes: Means and Standard Deviation are actual values from Table 2 from Hien et al., (2004); Correlations are assumed and not based on actual data.

**Table 2**

Population Parameters Derived from Table 1 Sufficient Statistics

<b>Effects</b>	<b>Parameters</b>
Mean of SS/WHE Dummy Variable	.500
Variance of SS/WHE Dummy Indicator	.250
Fixed Effects	
In-Treatment Slope (ITS)	.178
Post-Treatment Slope (PTS)	.138
Post-Treatment Intercept (PTI)	-.075
SS/WHE differences in ITS	-.209
SS/WHE differences in PTS	.110
SS/WHE differences in PTI	-.261
Individual-Level Random Effects	
In-Treatment Slope (ITS)	.201
Post-Treatment Slope (PTS)	.058
Post-Treatment Intercept (PTI)	.222
Within-Individual Residual <sup>a</sup>	.204 <sup>b</sup>
Site-Level Random Effects	
In-Treatment Slope (ITS)	.0041
Post-Treatment Slope (PTS)	.0012
Post-Treatment Intercept (PTI)	.0045

Notes.

<sup>a</sup>Constrained to equality across all timepoints.

<sup>b</sup>Average of the treatment condition-specific values. ITS = in-treatment slope. PTS = post-treatment slope. PTI = post-treatment intercept.

Table 3

## Statistical Power Estimates

Method	N	WA $\beta_1$	Avg. Est. $\beta_1$	St. Bias $\beta_1$	Coverage $\beta_1$	C $\beta_1$ Power	D $\beta_1$ Power	E $\beta_1$ Power	WA $\beta_1$ Power
LCPMM	150	-.214	-.2312	.1071	.920	.113	.169	.113	.382
	250	-.214	-.2266	.1057	.937	.336	.076	.346	.539
	353	-.214	-.2337	.1916	.926	.460	.085	.093	.678
	450	-.214	-.2296	.1710	.926	.546	.093	.100	.755
LGM	150	-.214	-.2311	.1196	.943	-	-	-	.411
	250	-.214	-.2270	.1124	.925	-	-	-	.589
	353	-.214	-.2352	.2158	.916	-	-	-	.737
	450	-.214	-.2307	.2026	.919	-	-	-	.799

Notes: LCPMM = Latent Class Pattern Mixture Model, LGM = Latent Growth Model, WA  $\beta_1$  = Weighted-averaged treatment effect on the in-treatment slope, C  $\beta_1$  = Completers treatment effect on the in-treatment slope (60% of the population), D  $\beta_1$  = Dropouts treatment effect on the in-treatment slope (20% of the population), E  $\beta_1$  = Erratics treatment effect on the in-treatment slope (20% of the population), Avg. Est. = Average Estimate across 1000 replications, St. Bias = Standardized Bias (values above .40 indicate problematic bias; Collins et al., 2001). Coverage = values below .90 indicate problematic coverage rates.



Table 4

## Supplemental Analysis Estimates

Method	N	WA $\beta_1$	Avg. Est. $\beta_1$	St. Bias  $\beta_1$	Coverage $\beta_1$
LCPMM	150	.0667	.0109	.3628	.918
	250	.0667	.0131	.4580	.923
	353	.0667	.0173	.5135	.926
	450	.0667	.0149	.5849	.901
LGM	150	.0667	-.0106	.5601	.911
	250	.0667	-.0125	.7266	.879
	353	.0667	-.0074	.8459	.878
	450	.0667	-.0107	.9462	.815

Notes. LCPMM = Latent Class Pattern Mixture Model. LGM = Latent Growth Model. WA  $\beta_1$  = Weighted-averaged treatment effect on the in-treatment slope. Avg. Est. = Average Estimate across 1000 replications. St. Bias = Standardized Bias (values above .40) indicate problematic bias; Collins et al., 2001). Coverage = values below .90 indicate problematic coverage rates.