



Published in final edited form as:

Am J Drug Alcohol Abuse. 2011 September ; 37(5): 383–391. doi:10.3109/00952990.2011.600386.

Modeling site effects in the design and analysis of multisite trials

Daniel J. Feaster, Ph.D.¹, Susan Mikulich-Gilbertson, Ph.D.², and Ahnalee M. Brincks, M.A.¹

¹Biostatistics Division, Department of Epidemiology and Public Health, University of Miami Miller School of Medicine

²Departments of Psychiatry and Biometrics/Informatics, University of Colorado, Denver

More than half of the published trials from NIDA's Drug Abuse Treatment Clinical Trials Network (CTN) have demonstrated significant site or site-by-treatment effects (1). This fact alone is testament to the importance of carefully considering the inclusion of sites in the analysis of multisite clinical trials of drug abuse treatment. The statistical choices for modeling site effects on treatment outcomes have important implications for both trial planning and interpretation of findings. We present and discuss different options for handling site effects in the statistical model and the impact of each on the construction and interpretation of the treatment estimate.

A researcher's perspective on site effects should be at least partly influenced by the aims of the project. If the project is focused more on establishing the efficacy of a treatment, and multiple sites are used primarily to facilitate subject recruitment, the influence of site can be viewed as a nuisance in the statistical analysis (2). Most pharmaceutical trials, particularly those for relatively rare conditions, fit into this category. In these trials, site effects are considered variance due to methods that need to be controlled in the analyses. Conversely, if the goal of the research is to establish that a particular treatment approach has broad applicability, emphasizing more the effectiveness side of the research continuum, then the influence of site becomes significantly more important and variability across sites may provide insight into the conditions under which the treatment approach is effective.

There are multiple factors that might cause there to be site differences in a research study. The effectiveness of a treatment approach may be related to some characteristic(s) of the individuals that tend to be found at a particular site. For instance, it is not unreasonable to think that participants within a site might have more in common with one another than with participants from other sites. Think of common observable characteristics that might be correlated with the outcome. The characteristics that cause this homogeneity might come from the clients themselves. For example, if socioeconomic status (SES) is related to the outcome, and there are large differences in SES among the sites, analyses will show site effects on the outcome variable as a result of these differences in SES. The effectiveness of a treatment approach might also be the result of a contextual aspect of the site. For example, a site may only serve participants whose primary problem is alcohol, or opioids, or some other drug category. Sites might also have particular treatment approaches/philosophies, or payment/reimbursement systems, which influence the type of client drawn to that site. Such site characteristics can cause the client pool within a site to be more homogenous than clients across multiple sites. Finally, there may be site-related factors that lead to differential treatment response. Particular treatment approaches/philosophies or payment/reimbursement

Declaration of Interest: The authors have no declaration of interests associated with this manuscript. This research was conducted as a part of the National Institutes on Drug Abuse' Clinical Trials Network and was supported by U10-DA13720.

systems may not only draw a particular client population, but may also directly influence the treatment outcomes. For example, there may be management or personnel differences across sites that result in differences in treatment response. Finally, implementation of the treatment conditions might unintentionally vary across sites.

Statistical approaches to managing site differences

There are three broad analytic approaches to managing site effects: 1) ignore them, 2) model them as fixed effects, and 3) model them as random effects. As we will show, if site effects are present, ignoring them can have deleterious consequences on the quality of the statistical inference. In this case, modeling site effects as either fixed or random may be the more appropriate choice. As we will discuss, in the case of continuous outcomes and linear models (i.e. models that are linear in the parameters), many of the differences between these approaches disappear when sites have equal sample sizes and equal proportions of participants assigned to each treatment condition. Unfortunately, most trials do not have equal sample sizes across sites. The more these sample sizes differ, the larger the difference in treatment effect estimates across the statistical methods for handling site.

We present the impact of modeling site effects using these three approaches in models including either a main effect of site alone, or the main effect of site and the site by treatment interaction. Throughout, we assume that randomization occurs within sites (i.e., individuals, not sites, are the level of randomization), that the outcome variable is continuous, and the statistical model is linear. We utilize an analysis of covariance approach to describe how various estimators of treatment effects work in the multi-site framework:

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij} \quad (1.1)$$

where $i=1, \dots, n_j$ indexes individuals within site, $j=1, \dots, S$, indexes sites, x_{ij} is a dichotomous treatment indicator and y_{ij} is the outcome for subject i in site j and e_{ij} is the residual error. Both y_{ij} and x_{ij} can be decomposed to their within-site and between-site variation. These distinct sources of variation could be utilized in two simple linear models to obtain two different estimates of the treatment effect. The within site estimate removes the site-specific mean of the outcome and treatment assignment from each individual's value of the respective quantity:

$$y_{ij} - \bar{y}_{.j} = \beta_w (x_{ij} - \bar{x}_{.j}) + (e_{ij} - \bar{e}_{.j}) \quad (1.2)$$

and the between site estimate is based on the site-specific means of the outcome and treatment assignment :

$$\bar{y}_{.j} = \alpha + \beta_b \bar{x}_{.j} + \bar{e}_{.j} \quad (1.3)$$

The within and between estimates are based on an orthogonal decomposition and are therefore independent sources of variation. These two estimates of the treatment effect are used later in our examination of fixed and random site effects.

In our presentation, we first discuss the implications of ignoring site in the analysis. Next, we consider the model with only a main effect of site and then the model that also includes the site-by-treatment interaction; for both of these models, the ramifications of specifying site as fixed or random are discussed. A brief example of the sample size requirements for these different models is presented and we conclude with a discussion and summation of the issues in addressing site effects for multisite trials.

Ignoring site effects

Kraemer (3) illustrated that in the presence of mean differences in outcome across sites, ignoring the influence of site can lead to widely varying estimates for the relationship between treatment condition and outcome. These estimates differ depending on the proportion of subjects at each site and whether the number of subjects assigned to the treatment conditions varies across sites.

Ignoring site may also bias the estimate of the standard error of the treatment effect if there is a non-trivial tendency for participants within a site to be more similar on the outcome variable than participants between sites. Ignoring this homogeneity in the analyses will lead to an overstatement of the statistical significance of the treatment effect (4, 5) even when participants are randomized within site (6). The magnitude of this bias depends both on how much of the variance in outcome is explained by site (the intra-class correlation, ICC) and the number of participants recruited at each site (7, 8, 9). In most trials within the CTN, the number of participants per site is at least fifty and frequently closer to 100. Even if site only explains 1% of the variance in outcome ($ICC=.01$), this leads to an inflation of variance by 25% or 50% for site sizes of 50 and 100, respectively. This, in turn implies that standard errors estimated ignoring this site effect would need to be inflated by 12% and 22%, respectively to be accurate. Thus, because it is nearly impossible to rule out such small site effects prior to running a multisite trial, ignoring the influence of site is really not a viable option.

Main effect of site

Fixed Main Effect for Site

The influence of site on the outcome may be estimated as a fixed effect by including an indicator variable for each site:

$$y_{ij} = \alpha + \beta_{f1} x_{ij} + a_j d_j + e_{ij}, \quad \sum_{j=1}^S a_j = 0, \quad e_{ij} \sim N(0, \sigma_e^2) \quad (1.4)$$

where d_j is a dummy indicator for site j , α is the outcome mean for the control group, β_{f1} is the difference between the control mean (α) and treatment group mean (i.e., the treatment effect), a_j is the deviation of site j mean outcome from the overall mean of outcome. The restriction allows the inclusion of dummy indicators for all sites and a constant term in the full-rank model. This approach results in a within-site analysis of treatment on outcome because inclusion of site-specific dummy variables causes each predictor (e.g. treatment) to be deviated from its site-specific mean. Thus the treatment estimate in this case is based on within-site variation ($\beta_{f1} = \beta_w$, from Equation 1.2.) and the between-site variation is removed from the treatment estimate. When there is only a fixed main effect for site (i.e. when a treatment-by-site interaction is not included in the model) the estimate of treatment effect based on the standard Type III sums of squares is weighted by the sample size of the site. As a result, if there is wide variation in sample size across sites, this variation is incorporated into the analysis and the estimated treatment effect is the weighted mean of treatment effects across sites.

In a study with equal proportion of individuals across sites, and an equal number assigned to each treatment condition, the between-site variation in treatment assignment cannot explain any of the variation in the outcome variable. To illustrate this, consider an estimate of treatment effect based on between-site variability (Equation 3). The between-site treatment indicator, $\bar{x}_{.j}$ is the proportion of participants assigned to the experimental treatment site j . If

sample size and the proportion assigned to the experimental treatment is the same at each site, this average of the treatment indicator will not vary across sites, therefore treatment assignment cannot explain any of the variability in the average outcome across sites. Thus, when we have consistent proportions of individuals assigned to treatment condition across site and there is no site-by-treatment interaction, a within-site analysis of treatment effects does not involve any loss of information.

Examples of Trials Using Fixed Effects for Site

The first two trials within the CTN to test the effectiveness of contingency management + treatment as usual (TAU) versus TAU alone for stimulant abuse analyzed their data with and without a fixed site effect in the model. CTN 0006 involved eight outpatient psychosocial treatment programs (10). Including site as a predictor of outcome did not alter conclusions of the trial (that contingency management was effective in increasing drug-free urines, increasing retention and duration of abstinence). Similarly, in CTN 0007, which included six methadone maintenance sites, Site was an important predictor in supplemental analyses, but did not alter the conclusions that contingency management effectively increased stimulant abstinence in community-based methadone maintenance treatment clinic sites (11).

Advantages and Disadvantages of Fixed Main Effect for Site

The primary advantage of using a fixed site effect approach is ease of implementation. Another advantage is that When the number of participants per site is equal and assignment to condition is balanced, sample size determination is straight forward. The only factor influencing the number of sites in the study is choosing enough sites to reach the necessary overall sample size for adequate power. The total sample size needed is the same regardless of whether all participants come from a single site, or from multiple sites (though an additional participant should be included per-site to account for the degrees of freedom used in estimating the site-specific means). However, if a study plans for equal sample sizes across sites, but in reality achieves unbalanced sample sizes across sites, there is a reduction in statistical power (12).

The primary disadvantage of the fixed site effect approach is that the analysis is conditional on the sites within the model, i.e. there is no statistical basis for generalization to the larger population of sites that could have been in the study, or for whom the intervention might eventually be implemented. Just as in a single-site study, the only way to assess whether the treatment is likely to have similar results with a particular type of patient is to qualitatively evaluate whether the sample in the study was similar to that patient type. The study gives no information on how to statistically assess the likelihood of treatment effectiveness for a participant from a different site. Another disadvantage is that the interpretation of an overall site effect depends on the assumption of no site-by-treatment interaction (see below).

Random Main Effect for Site

In a random site effect model a variance component is included in the model to account for variability in the outcome due to differences between sites:

$$y_{ij} = \alpha + \beta_{r1} x_{ij} + s_j + e_{ij}, \quad s_j \sim N(0, \sigma_s^2), \quad e_{ij} \sim N(0, \sigma_e^2) \quad (1.5)$$

where s_j is the random effect for site j , and is normally distributed with mean 0 and variance σ_s^2 , independently from the residual errors, e_{ij} . The generalized least squares estimate for β_{r1} , the treatment effect, is based on a weighted average of the within-site and between-site component estimates of treatment effect on outcome (Maddala, 1977):

$$\widehat{\beta}_{r1} = \left[\frac{\sum_{j=1}^s \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2}{\sum_{j=1}^s \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 + \theta \sum_{j=1}^s (\bar{x}_{.j} - \bar{x}_{..})^2} \right] b_w + \left[\frac{\theta \sum_{j=1}^s (\bar{x}_{.j} - \bar{x}_{..})^2}{\sum_{j=1}^s \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 + \theta \sum_{j=1}^s (\bar{x}_{.j} - \bar{x}_{..})^2} \right] b_b, \quad (1.6)$$

where $\theta = \frac{\sigma_e^2}{\sigma_e^2 + S\sigma_s^2}$ and the weights are the quantities in large square brackets. Note that the weight on b_w has the within-site variation in treatment assignment in the numerator whereas the weight on b_b has the between-site variation in treatment assignment in the numerator. If neither the sample sizes nor the proportions assigned to the treatment conditions differ across sites then the between-site variability in treatment assignment is zero (causing $\bar{x}_{.j} - \bar{x}_{..} = 0$ for all j , the weight on the within-site component, b_w to be 1 and the weight on the between-site component b_b to be 0). In this case, both the estimate and standard error for the treatment effect will equal that obtained from the fixed site effect model (13). However, if sample sizes or proportions of participants in the treatment condition vary across sites then the treatment estimate will be a weighted average of the within-site and between site components of treatment differences. Since the random site effect estimate incorporates more information (i.e., the between-sites component) than the fixed site effect estimate, the standard errors of the treatment effect should be smaller in the random effects model.

Examples of Trials Using Random Main Effect for Site

There have been three CTN trials, thus far, in which site was modeled as a random effect: The five-site CTN-0005 which tested motivational interviewing versus TAU (14), the seven-site CTN-0009 which tested smoking cessation treatment +TAU versus TAU alone (15), and the 12-site CTN 0019 which tested safer sex skills building groups versus standard HIV/STD educational groups for women in reducing HIV sexual risk behavior (16). As is frequently the case when these models are reported in peer-reviewed journals, no information on the variance associated with site was reported in these three trials. The CTN-0005 trial did present treatment effects within each site for the major outcomes and discussed site variability in the context of a fixed-effect estimate for the trial's treatment fidelity measure.

Advantages and Disadvantages of Random Main Effect for Site

The primary advantage of the random site effect approach is that the treatment effect estimate is always as efficient (i.e., has at least as small a standard error) as the estimate produced by a fixed site effect model. When there are imbalances in sample size or proportions of individuals in treatment conditions across sites, the random site effect approach will produce a more efficient estimate for treatment effect than that of a fixed site effect approach. The disadvantage of this approach is that it requires a minimum number of sites. Some have endorsed a minimum of five sites as a guideline (13). Although some recommend a higher number of sites, these recommendations are for the case where there is an attempt to model a site-level covariate as an explanation of site variability (17). Unlike its fixed effect counterpart, the random site effect model does provide a statistical basis for generalizing intervention results to sites outside of the trial. The representativeness of the sites participating in the trials should be taken into consideration when seeking to generalize results to sites outside the trial. Any generalizations beyond the trial sites is only appropriate with the strong assumption that there is not a site-by-treatment interaction.

Site-by-treatment interaction

If sites differ in the effectiveness of the experimental treatment then there is a site-by-treatment interaction. Just as it is impossible to predict that there will not be a main effect of site, it is also impossible to predict that a multisite trial will not have a site-by-treatment interaction. Kraemer and Robinson (18) strongly advocate for the inclusion of both site and site-by-treatment interaction in analyses of multisite clinical trials. The findings from many studies within the CTN also support the need to consider the site-by-treatment interaction (14; 19; 20).

How the main effect for site is specified in the model determines how the site-by treatment interaction is specified, i.e., if site is considered a fixed (random) effect then the interaction is considered a fixed (random) effect. The two methods of modeling the site-by-treatment interaction, as either a fixed or random effect, have very different underlying philosophies. The random effect approach assumes there is variability in the treatment effect across sites and tries to estimate this variability so that the resulting estimate of the treatment effect tests whether, *on average*, the experimental treatment is better than the control treatment. The fixed-effect approach attempts to control for both the main effect of site and the site-by-treatment interaction, but does not model the variability in these terms resulting from site differences.

The fixed effect method is generally undertaken with the hope that the site-by-treatment interaction will be statistically insignificant and can be dropped from the model. However, the the site-by-treatment interaction is frequently underpowered, implying that a statistically non-significant site-by-treatment interaction may have little meaning (18). To compensate, the site-by-treatment interaction is frequently evaluated using a higher alpha level (e.g. $p < .10$; 21). When there is a significant site-by-treatment interaction in the fixed effect approach, there is disagreement on whether to report the overall treatment effect and if so, how to calculate this estimate. Technically, a statistically significant site-by-treatment interaction means that the effect of treatment should be analyzed and reported separately for each site. However, particularly in pharmaceutical trials, a great deal of attention is paid to how to generate an overall estimate of the treatment effect in the presence of a significant site-by-treatment interaction (22; 23).

Fixed Effects for Site and Site-By-Treatment Interaction

The model with all sites included as predictors in addition to each of the site-by-treatment interactions might look like:

$$y_{ij} = \alpha + \beta_{f2}x_{ij} + a_j d_j + b_j d_j x_{ij} + e_{ij}, \quad \sum_{i=1}^s a_j = 0, \quad \sum_{i=1}^s b_j = 0, \quad e_{ij} \sim N(0, \sigma_e^2) \quad (1.7)$$

where α and β_{f2} are as in the fixed site effect model (Equation 1.4). Here a_j is the difference between the control group mean from site j and the overall control group mean (α). The site-by-treatment interaction b_j is the difference between the treatment group mean for site j and the overall treatment effect (β_{f2}). Thus, β_{f2} is a simple average of the treatment effect estimates for each site (22). The two restrictions are necessary to identify deviations for all sites and the α and β_{f2} .

In this analysis, each site is counted equally, regardless of the number of participants within the site when standard sums of squares are used (Type III SS). However, some argue that Type II sums of squares when sample sizes at each site differ so that estimates of the treatment effect will be weighted by the sample size within each site (24; 25). Type II SS

control for all variables in the model except for any interactions with the effect of interest, here treatment.

Examples of Trials Using Fixed Effects for Site and Site-by-Treatment

Five CTN trials have utilized fixed effects for site and the site by treatment interaction. All three of the trials for motivational enhancement therapy (MET) used this strategy to test the effect of MET versus TAU: The original five-site MET in community-based drug abuse clinics (CTN-0004; 19), the four-site MET for pregnant women (CTN-0013; 20), and the five-site MET for Spanish-speaking substance users (CTN-0021; 26). In two of these trials (19, 20) there were marginally significant site-by-treatment interaction effects, highlighting the importance of adequate power for these interaction effects. In all three MET trials treatment effects within each site were fully explored despite the absence of statistical significance for the site-by-treatment interaction. In the other two studies, the six-site CTN-0010, comparing a 12-week treatment versus a 14-day detox with Buprenorphine-Naloxone for opioid-addicted youth (27), and the seven-site CTN-0015 Seeking Safety intervention for PTSD + TAU versus women's health education + TAU (28), the site-by-treatment interaction was found to be non-significant and subsequently dropped from the final model.

Advantages and Disadvantages to Fixed Effects for Site and Site-By-Treatment

The advantages and disadvantages of this fixed effect model with site and site-by-treatment interaction are similar to those with only fixed main effects. Advantages include ease of implementation and no need for a minimum number of sites. The primary disadvantage is again that the analysis is conditional on the sites within the model, with resulting implications for generalizability. There is also the question of whether treatment effects should be combined across sites to create an overall estimate of treatment effect if there is a clinically important site-by-treatment interaction. Within the Clinical Trials Network, concern about the existence of an important site-by-treatment interaction has led some trials to recruit sufficient samples at each site to show significant treatment effects within sites (29). When treatment effects are combined across sites (in the presence of a statistically insignificant site-by-treatment interaction, for instance), an additional disadvantage is that the standard Type III SS approach weights site-specific treatment effects equally. For studies with very different sample sizes across sites and/or different proportions of treatment assignment within site, this may be undesirable. This can be "corrected" by using Type II SS, however, this is similar to using the main effects model in the presence of significant site-by-treatment interactions. In addition, statistical power in both the unweighted analysis (Type III SS) approach and in the weighted (Type II SS) approach is reduced in studies with large variability in site sample sizes or proportion of treatment assignment within sites (30; 12).

Random Effects for Site and Site-By-Treatment

The model including random effects for both site and site-by-treatment could be represented as:

$$y_{ij} = \alpha + \beta_1 x_{ij} + s_{1j} + x_{ij} s_{2j} + e_{ij}, \quad sN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{s00}^2 & \sigma_{s01}^2 \\ \sigma_{s01}^2 & \sigma_{s11}^2 \end{bmatrix} \right), e_{ij} \sim N(0, \sigma_e^2) \quad (1.8)$$

Here s_{1j} is the random effect associated with site and s_{2j} is the random effect associated with the site-by-treatment interaction (recall x_{ij} is the treatment indicator). This model is similar to that utilized in Raudenbush and Liu (31). As in the simple random site effect model, the estimate of the treatment effect includes information from both within- and between-sites

and is a precision weighted average of the site-specific treatment effects. These weights incorporate the respective site sample sizes because the precision is the inverse of the variance. The standard error of the treatment effect includes the variance associated with the site-by-treatment interactions (31). If the site-by-treatment variance is not zero, the standard error of the treatment effect is larger in this random site effect model (with interaction) than either the fixed-effect model (with interaction) or the random site effect model (without interaction). This means a larger sample size is required to achieve statistical significance (see 32 for an empirical example).

Examples of Trials Using Random Effects for both Site and Site-by-Treatment

There have been two trials in the CTN in which random effects for both site and site-by-treatment were included. The eight-site CTN-0014 tested Brief Strategic Family Therapy for adolescent drug abuse versus TAU and examined trajectories of drug use over 12 months (33, 34). The planned analysis included random effects for the trajectory components, site and site by treatment interactions. Similarly, the eight-site CTN-0017, which tested a counseling and education intervention for HIV risk behaviors versus a therapeutic alliance intervention + treatment as usual versus treatment as usual (35), also examined trajectories (of HIV risk) in which site and its interaction with other effects such as site-by-treatment were modeled as random effects. In both of these trials, the variance associated with the site-by-treatment interaction was found to be zero and the random effect was dropped from the final models.

Advantages and Disadvantages of Random Effects for Site and Site-By-Treatment

The primary advantage of this approach is that there are clear statistical grounds for generalizing the results to participants treated in sites that were not part of the study. Another advantage is that there is no ambiguity about how to calculate the overall treatment effect in this model even when there is variability in treatment effects across sites. Finally, this approach makes it possible to include characteristics of the site as predictors of site and site-by-treatment effects. For example, consider a trial where sites could be characterized into subtypes (such as residential or outpatient). In a statistical model with random effects for both site and site-by-treatment interaction, it would be possible to include site subtype and the subtype-by-treatment interaction as covariates with the two site-related random effects representing site variability in excess of that explained by site-type. In the fixed-effects approach, however, the fixed effects of site and site-by-treatment interaction estimate the mean differences across site and thus control for all the variability associated with these factors (17). Therefore, there is no remaining variability for site characteristics (e.g. subtype) to predict if they were to be included in the fixed effects model. The random-effects model provides a natural way to incorporate these predictors enabling the researcher to assess how much of the site variability these factors explain (36).

The disadvantage of the random-effects approach is that more sites are required to provide stable estimates of the variance terms. In a simple post-test model (perhaps controlling for baseline values of the outcome variable) this may not be so problematic because as few as five sites may be sufficient (13; but see example in Table 1, described later). In repeated measures designs, random effects are frequently included to model the non-independence of repeated measures within a subject. Inclusion of random effects for both the repeated measures and for the site and site-by-treatment interactions generally require more sites in these models, though more research is needed in this area. It may be possible to incorporate prior information on site variability (from other trials, for example) using a Bayesian approach to reduce the number of sites needed for stable estimates (37; 38; 39).

Statistical power considerations

Statistical power for clinical trials using the fixed site effects models can be estimated using standard methods (e.g., 40). There are fewer tools for statistical power calculations for trials using random site effect models. Raudenbush and Liu (31) demonstrated how to calculate power for a random site effects model with a simple post-test analysis (which could control for baseline values). The W.T. Grant Foundation funded the development of an accompanying computer program, Optimal Design 2.0 (41). To provide an illustration of the power differences between the random and fixed site effects approaches we conducted a comparative power analysis using this program.

Our goal was to examine the influence of sample size (S = number of sites, n = number of individuals per site, N = total sample size) on statistical power. For the fixed site effect approach we examined the sample size needed within sites to have 1) adequate power for a treatment effect and 2) power for the site-by-treatment interaction. For the random site effect approach we examined the sample size needed within sites (n) to have 1) adequate power for a treatment effect and 2) adequate power to show that a site-level covariate other than treatment condition significantly influenced treatment outcomes. All cases assume a moderate effect size ($d=0.50$) for the average treatment effect and that the site-by-treatment interaction explained 5.9% of the variability in the outcome variable (equivalent to a moderate effect size for the interaction, 40). The effect size for the site-level covariate was also fixed at $d = 0.50$. For a single site study, with $d=0.5$, 128 participants are needed to achieve 80% power. In a multi-site study including a fixed-effect for site, the sample required for 80% power for the overall treatment effect would equal $N' = N + S$, i.e. 128 plus the number of sites (to account for the degrees of freedom used in estimating separate site means).

Table 1 shows the resulting sample size requirements. For the smallest number of sites, six, the fixed effect model with the interaction has the smallest overall sample size ($N = 300$). In the random effect approach, N falls rapidly as the number of sites (S) increases. N also decreases quickly for the fixed site effect model with interactions, though not as quickly as for the random site effect model. The net effect is that the N is lowest at $S = 8$ or more sites in the random site effect model. Obtaining adequate power for a treatment effect at each site requires the largest sample size. Recall that 128 participants are required for 80% power in a single site study. This is also the sample size required per site (n), if the goal is to obtain 80% power within each site. Finally, the column for site covariates illustrates the importance of a larger number of sites when the aim is to explain the site-level variability. At the smaller numbers of sites it is not possible to achieve 80% power when $d=.50$.

Recommendations for researchers

Ignoring site is not a viable option in multi-site clinical trials. Omitting site effects in the analyses can bias the treatment effect and bias the standard error associated with the treatment effect if participants within sites are more homogeneous than participants between sites (3; 4; 5; 6). In addition, within site randomization means that randomization is stratified by site. Most statisticians would recommend that when randomization is stratified by a set of characteristics, that these characteristics should be included in the analysis (42, 43).

The distinction between efficacy trials and effectiveness trials is rarely sharp; the two types of trials are two extremes of a continuum (44; 3). Most trials, other than the first tests of a new intervention, have elements of both. To the extent that a trial is closer to the efficacy side of the spectrum, specifying fixed site effects for site is a logical approach because the focus is on establishing that the treatment can be efficacious rather than on generalizability. As a trial moves toward the effectiveness side of the continuum, specifying random site

effects may be beneficial because the goal includes the ability to generalize the efficacy of a treatment to new populations and clinics. In all cases, the advantages and limitations of the chosen methods for handling site effects should be made explicit in the presentation of results.

Whereas a fixed site effect approach provides no statistical grounds for generalizing to sites beyond those included in the trial, other approaches to generalization can be taken. Even in a single site trial, participants are rarely randomly sampled, yet the results are frequently generalized, at least to the types of participants that were included in the trial. Edgington (45) labels this nonstatistical inference, that is, inference without a basis in probability, and refers to the resulting nonstatistical generalization as a standard scientific procedure. For example, when assessing the results of a single-site trial, providers might look at qualitative characteristics of the study's sample and characteristics of the setting in which the research was based to assess whether the intervention approach might work well for their situation. A similar approach can be used in a multisite trial. This requires a good, comprehensive description of the sites involved in the trial so that the types of sites to which the research can be generalized is clear. However, no statistical inference is involved in this process.

Within a fixed-effect approach it would be important to either power the study to uncover a significant site-by-treatment interaction or power the study to show significant effects within sites (as was done for several of the early CTN trials; 29). Whereas the latter approach may not show significant differences between sites, it should have adequate power to show treatment effects within those sites with the larger effect sizes. However, if the researcher is interested in making statements about what type of sites are most likely to be successful implementing a particular treatment or what characteristics of sites are associated with better outcomes, then a random-effects approach is needed and the number of sites required and larger sample sizes within sites become important. The number of sites required to have sufficient power for a covariate effect is considerably larger than the number required for the overall treatment effect. Most trials within the CTN have not involved this many sites. However, it may be possible to examine the effect of sites on treatment by combining information across trials. A more standardized approach to collection of information about the sites would facilitate both attempts to characterize how trial results may be generalized to other sites (regardless of the statistical methods used within a trial) and the impact of site characteristics across trials.

Acknowledgments

This research was conducted as a part of the National Institutes on Drug Abuse' Clinical Trials Network and was supported by U10-DA13720.

References

1. Nunes EV, Ball S, Booth R, Brigham G, Calsyn DA, Carroll K, Woody G. Multisite effectiveness trials of treatments for substance abuse and co-occurring problems: Have we chosen the best designs? *Journal of Substance Abuse Treatment*. 2010; 38(Supplement 1):S97–S112. [PubMed: 20307801]
2. Lewis JA. Discussion on the paper by Senn. *Journal of the Royal Statistical Society Series D*. 2000; 49:156–157.
3. Kraemer HC. Pitfalls of multisite randomized clinical trials of efficacy and effectiveness. *Schizophrenia Bulletin*. 2000; 26(3):533–541. [PubMed: 10993394]
4. Locallo AR, Berlin JA, Ten TR, Kimmel SE. Adjustments for center in multicenter studies: An overview. *Annals of Internal Medicine*. 2001; 135:112–123. [PubMed: 11453711]
5. Wampold BE, Serlin RC. The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*. 2000; 5(4):425–433. [PubMed: 11194206]

6. Lee K, Thompson SG. Clustering by health professional in individually randomised trials. *British Medical Journal*. 2005; 330(7483):142–144. [PubMed: 15649931]
7. Hutchison D. Designing your sample efficiently: clustering effects in education surveys. *Educational Research*. 2009; 51(1):109–126.
8. McCoach DB, Adelson JL. Dealing with dependence (Part I): Understanding the effects of clustered data. *Gifted Child Quarterly*. 2010; 54:152–155.
9. Kenny DA. The effect of nonindependence on significant testing in dyadic research. *Personal Relationships*. 1995; 2:67–75.
10. Petry NM, Peirce JM, Stitzer ML, Blaine J, Roll JM, Cohen A, Li R. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: A national drug abuse treatment clinical trials network study. *Archives of General Psychiatry*. 2005; 62:1148–1156. [PubMed: 16203960]
11. Peirce JM, Petry NM, Stitzer ML, Blaine J, Kellogg S, Satterfield F, Li R. Effects of lower-cost incentives on stimulant abstinence in methadone maintenance treatment: A National Drug Abuse Treatment Clinical Trials Network study. *Archives of General Psychiatry*. 2006; 63(2):201–208. [PubMed: 16461864]
12. Ruvuna F. Unequal center sizes, sample size, and power in multicenter clinical trials. *Drug Information Journal*. 2004; 38(4):387–394.
13. Brown, H.; Prescott, R. *Applied Mixed Models in Medicine*. Wiley; New York, NY: 1999.
14. Carroll KM, Ball SA, Nich C, Martino S, Frankforter TL, Farentinos C, Woody GE. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: A multisite effectiveness study. *Drug and Alcohol Dependence*. 2006; 81:301–312. [PubMed: 16169159]
15. Reid MA, Fallon B, Sonne S, Flammio F, Nunes EV, Jaing H, Rotrosen J. Smoking cessation treatment in community-based substance abuse rehabilitation programs. *Journal of Substance Abuse Treatment*. 2008; 35:68–77. [PubMed: 17951021]
16. Tross S, Campbell ANC, Cohen LR, Calsyn D, Pavlicova M, Miele GM, Hu M, Haynes L, Nugent N, Gan W, Hatch-Maillette M, Mandler R, McLaughlin P, El-Bassel N, Crits-Christoph P, Nunes EV. Effectiveness of HIV/STD sexual risk reduction groups for women in substance abuse treatment programs: Results of NIDA clinical trials network trial. *Journal of Acquired Immune Deficiency Syndrome*. 2008; 48(5):581–589.
17. Snijders, T.; Bosker, R. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. SAGE Publications; London: 2000.
18. Kraemer HC, Robinson TN. Are certain multicenter randomized clinical trial structures misleading clinical and policy decisions? *Contemporary Clinical Trials*. 2005; 26(5):518–529. [PubMed: 15993653]
19. Ball SA, Martino S, Nich C, Frankforter TL, Van Horn D, Crits-Christoph P, Carroll KM. Site matters: Multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology*. 2007; 75(4):556–567. [PubMed: 17663610]
20. Winhusen T, Kropp F, Babcock D, Hague D, Erickson SJ, Renz C, Somoza E. Motivational enhancement therapy to improve treatment utilization and outcome in pregnant substance users. *Journal of Substance Abuse Treatment*. 2008; 35(2):161–173. [PubMed: 18083322]
21. Kallen A. Treatment-by-center interaction: What is the issue? *Drug Information Journal*. 1997; 31:927–936.
22. Fleiss JL. Analysis of data from multiclinic trials. *Controlled Clinical Trials*. 1986; 7(4):267–275. [PubMed: 3802849]
23. Lin Z. An issue of statistical analysis in controlled multi-centre studies: How shall we weight the centres? *Statistics in Medicine*. 1999; 18(4):365–373. [PubMed: 10070679]
24. Senn S. Some controversies in planning and analysing multi-centre trials *Statistics in Medicine*. 1998; 17(15-16):1753–65.
25. Senn S. Consensus and controversy in pharmaceutical statistics. *Journal of the Royal Statistical Society*. 2000; 49(2):135–176. Series D (The Statistician)

26. Carroll KM, Martino S, Ball SA, Nich C, Frankforter T, Anez LM, Farentinos C. A multisite randomized effectiveness trial of motivational enhancement therapy for Spanish-speaking substance users. *Journal of Consulting and Clinical Psychology*. 2009; 77(5):993–999. [PubMed: 19803579]
27. Woody GE, Poole SA, Subramaniam G, Dugosh K, Bogenschutz M, Abbott P, Patkar A, Publicker M, McCain K, Potter JS, Forman R, Vetter V, McNicholas L, Blaine J, Lynch KG, Fudala P. Extended versus short-term buprenorphine-naloxone for treatment of opioid-addicted youth: A randomized trial. *Journal of the American Medical Association*. 2008; 300(17):2003–2011. [PubMed: 18984887]
28. Hien DA, Wells EA, Jiang H, Suarez-Morales L, Campbell ANC, Cohen LR, Miele GM, Killeen T, Brigham GS, Zhang Y. Multi-site randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. *Journal of Consulting and Clinical Psychology*. 2009; 77(4):607–619. [PubMed: 19634955]
29. Carroll KM, Farentinos C, Ball SA, Crits-Christoph P, Libby B, Morgenstern J, Woody GE. MET meets the real world: Design issues and clinical strategies in the Clinical Trials Network. *Journal of Substance Abuse Treatment*. 2002; 23:73–80. [PubMed: 12220604]
30. Lin Z. The number of centers in a multicenter clinical study: Effects on statistical power. *Drug Information Journal*. 2000; 34(2):379–386.
31. Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. *Psychological Methods*. 2000; 5(2):199–213. [PubMed: 10937329]
32. Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *Journal of Clinical Epidemiology*. 2003; 56(4):341–350. [PubMed: 12767411]
33. Feaster DJ, Robbins MS, Horigian V, Szapocznik J. Statistical issues in multi-site effectiveness trials: The case of brief strategic family therapy for adolescent drug abuse treatment. *Clinical Trials*. 2004; 1:428–439. [PubMed: 16279281]
34. Robbin MS, Szapocznik JS, Horigian VE, Feaster DJ, Puccinelli M, Jacobs P, Burlew K, Werstlein R, Bachrach K, Brigham G. Brief Strategic Family Therapy™ for adolescent drug abusers: A multi-site effectiveness study. *Contemporary Clinical Trials*. 2009; 30(3):269–278. [PubMed: 19470315]
35. Booth RE, Campbell BK, Mikulich-Gilbertson SK, Tillotson CJ, Choi D, Robinson J, McCarty D. Reducing HIV-related risk behaviors among injection drug users in residential detoxification. *AIDS and Behavior*. 2010; 15:30–44. [PubMed: 20652630]
36. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*. 1998; 23(4):323–355.
37. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1(3):515–533.
38. Gould AL. Multi-centre trial analysis revisited. *Statistics in Medicine*. 1998; 17(15-16):1779–1797. [PubMed: 9749447]
39. Subbiah M, Kumar BK, Srinivasan MR. Bayesian approach to multicentre sparse data. *Communications in Statistics - Simulation and Computation*. 2008; 37(4):687–696.
40. Cohen, J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates; Hillsdale, NJ: 1988.
41. Spybrook, J.; Raudenbush, SW.; Congdon, R.; Martinez, A. Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software. W.T. Grant Foundation. 2009. http://www.wtgrantfoundation.org/news/foundation_news/optimal-design
42. Friedman, LM.; Furberg, CD.; DeMets, DL. *Fundamentals of Clinical Trials*. Springer; New York, NY: 1998.
43. Rosenberger, WF.; Lachin, JM. *Randomization in Clinical Trials: Theory and Practice*. Wiley; New York: 2002.
44. Carroll KM, Rounsaville BJ. Bridging the gap: A hybrid model to link efficacy and effectiveness research in substance abuse treatment. *Psychiatric Services*. 2003; 54(3):333–339. [PubMed: 12610240]
45. Edgington, ES. *Randomization Test*. M. Dekker; New York: 1995.

Table 1

S n of Sites	Random Effect ¹			Fixed Effect ²		
	80% Power for Treatment Effect	80% Power for Site-Covariate	80% Power for Interaction	80% Power Within Each Site	Total N	Total N
	n per Site	n per Site	n per Site	n per Site	n per Site	n per Site
6	70	420	-	50	300	128
8	34	272	-	42	336	128
10	22	220	-	38	380	128
12	17	204	265	34	408	128
14	13	182	127	30	420	128
16	11	176	84	28	448	128
						2048

¹ Power calculated using Optimal Design Software: http://www.wtgrantfoundation.org/news/foundation_news/optimal-design. Power for the site-covariate calculated with the SAS-program from the appendix from Raudenbush and Liu (31).

² Power calculated using methods of Cohen (40)