

# THE NATIONAL TREATMENT IMPROVEMENT EVALUATION STUDY: MULTILEVEL REANALYSIS OF TREATMENT OUTCOMES<sup>1</sup>

Robert Orwin, Ronaldo Iachan, Bruce Ellis, Chuck Wolters  
Battelle Centers for Public Health Research and Evaluation  
2101 Wilson Blvd., Ste. 800, Arlington, VA 22201

**Key Words:** Multilevel Models, Substance Abuse Treatment, Program Evaluation

## 1.0 INTRODUCTION

Treatment outcome data from cross-site evaluations are inherently multilevel and hierarchical in that there exist multiple intervention sites, each with multiple clients. Statistically, clients are considered “nested” within sites. Often, the sites selected are formally sampled from a larger universe of sites to assure statistical representativeness. Other times they are not formally sampled but still function as a sample *conceptually*, that is, the scientific question is rarely limited to the effectiveness of specific programs at specific locations, but rather to a class of programs *like* them to permit generalization. Consequently, whether formally representative or just conceptually representative, cross-site or between-program effects are properly modeled statistically as random rather than fixed. Therefore, a multilevel, random-effects modeling approach is appropriate for analyzing these types of data.

In addition, multilevel models properly account for the nested structure of the data (e.g. individuals are embedded in treatment groups within different settings). Because of clustering within sites, measurements across individuals within sites may be correlated. Failure to account for this correlation--called the intraclass correlation or design effect-- and the additional variance attributable to random effects can result in underestimated standard errors and spurious statistical significance. Consequently, traditional statistical models that ignore the multilevel structure of the data are likely to produce erroneous results.

Finally, multilevel models more adequately address the direct effects of site characteristics on individual behaviors, and interactions between individual and site characteristics, than typically expressed in the framework of multiple regression and logistic regression. That is, terms can be specified as individual-level variables (level 1), site-level variables (level 2), or as interactions both within and across levels.

This paper describes a multilevel reanalysis of treatment outcomes from the National Treatment Improvement Evaluation Study (NTIES). The NTIES survey was a five-year longitudinal study of the impact of substance abuse treatment on a total of 5,388 clients purposively sampled from public substance abuse treatment programs (service delivery units or SDUs) that were funded by the Center for Substance Abuse Treatment (CSAT). Data were collected from five SDU modalities: methadone maintenance, non-methadone outpatient, short-term residential, long-term residential and correctional.

The original NTIES outcome analysis included multiple regressions of treatment outcomes at the client level, and a separate analysis of treatment effectiveness at the SDU level, adjusted for differences in case mix. Client-level analyses were conducted by performing multiple regressions within treatment modality on measures of improvement (pre-post change) in each of six severity domains: drug use, alcohol use, criminal activity, psychiatric problems, medical problems, and employment problems.

The case mix adjusted, SDU-level analysis was conducted in two stages. The first stage modeled client outcomes based on client characteristics, including pre-treatment severity. This produced a model-predicted outcome for each client, which was then averaged within SDUs (this procedure was performed for each of the six outcomes). The averaged value represented the predicted outcome for the SDU, based on the SDU's case mix. The second stage compared the actual aggregated outcomes for each SDU with the predicted outcome using SDU-level variables to model the comparison. SDU-level variables examined were: treatment modality, whether the SDU was selected for a Critical Populations grant, whether the SDU was selected for a Target Cities grant, average administrative, medical, and therapeutic costs per client per day, average length of stay and completion rate. The dependent variables were the predicted minus the actual average level of severity for each SDU in each outcome domain. Results showed that drug use severity was the only outcome measure that the SDU-level

---

<sup>1</sup> Produced under contract CSAT-97-7016 for the Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services. The opinions expressed herein are the views of the authors and do not necessarily reflect the official policy or position of the Federal Government.

variables predicted to a statistically significant degree, yet even on this outcome it explained only 3 percent of variance beyond the 18 percent already explained by case mix.

The original analysis had several limitations. First, the client-level multiple regression analyses were based on pooled cases across SDUs. Thus they did not account for the intraclass correlations or design effects caused by the clustering of clients within SDUs. While this did not affect overall estimates of treatment effectiveness, standard errors were biased downward and effect coefficients for client characteristics were more likely to appear statistically significant as a result.

Second, the SDU-level analysis was limited in the SDU characteristics it examined. Specifically, it examined modality, type of CSAT grant support, cost variables, and retention. It did not examine other variables that the literature suggests might also influence treatment effectiveness, such as whether the treatment program provided case management (Cox et al. 1993; Mejta et al. 1994; Orwin et al., 1994), matched clients to counselors or services (CSAT, 1994; 1995), tailored their program to specific populations (Westermeyer, 1990; Polinsky et al. 1998), and others. To some extent, the number of factors that could be examined was limited by the use of the SDU as the unit of analysis ( $n=71$ ).

Third, there are limitations inherent to the 2-stage approach, as noted in the report (NORC, 1997). A key one is the limited ability of the case-mix variables to reliably predict client outcomes in Stage 1, because most of the outcome variance remains unaccounted for. This in turn affects the average SDU outcome, which serves as the dependent variable in Stage 2.

Fourth, because the client-level and SDU-level analyses were conducted separately, it was not possible to examine interactions between SDU- and client-level factors, e.g., given an individual's pre-treatment characteristics profile, what program characteristics offer the best chance of effective treatment for that individual.

The present re-analysis addresses all these issues. First, clients are nested within SDUs so that intraclass correlations are properly accounted for in the computation of standard errors and effect sizes. Second, a broader range of SDU-level variables are examined, and examined separately by modality as described below. Third, the multilevel, random-effects approach is not dependent on how well client demographics and pre-treatment severity predict outcomes. Fourth, because the client- and SDU-level characteristics are analyzed together in a single multilevel model, interactions between SDU- and client-level factors can be tested.

Analysis questions were:

- What SDU-level factors appear to improve

client outcomes, and how do these relationships vary by modality?

- What interactions occur between client-level -and SDU-level factors?
- To what degree does the inclusion of SDU-level factors improve overall model fit?
- What are the implications for further analysis, policy, and practice?

## 2.0 METHODS

Outcome variables and client-level predictors (level 1 variables) were the same as those used in the original analysis (NORC, 1997), with some slight modifications (see Orwin et al., 1999a, for details). The choice of SDU-level variables was guided by two primary criteria: 1) the amount of nonmissing data in the variable, and 2) the variability of the variable across SDUs. The first criterion excluded variables whose values were missing for high numbers of observations. The second criterion excluded variables that would have low explanatory power due to their flat distribution.

We illustrate the model with a simple example using one single client-level variable,  $X$ , and no interactions. The notation gets more complex once we move to the more general cases considered in our analyses. With a formulation analogous to that presented in Singer (1999), for example, the model expresses the outcome  $Y_{ij}$  for client  $i$  in SDU  $j$  as a function of a vector  $Z$  of SDU-level variables:

$$Y_{i,j} = a_j + b_j X_{i,j} + e_i$$

$$a_j = c + \sum_k \{a_{j,k} Z_{j,k}\}$$

$$b_j = C + \sum_k \{g_{j,k} Z_{j,k}\}$$

where  $a_j$  is a random intercept for SDU  $j$ , and  $b_j$  is a random slope coefficient that relates the covariate  $X$  to the outcome  $Y$ . The latter two equations connect the random coefficients,  $a$  and  $b$ , to the SDU-level variables,  $Z$ . The error structure specifies that  $e_{ij}$  is normally distributed with mean 0 and standard deviation  $s$  and  $E_j, F_j$  have a joint bivariate normal distribution with mean 0.

This approach allows the introduction first of client-level variables for each modality, then SDU-level. With the addition of SDU-level variables for each modality, we can assess the improvement in the model fits. For this purpose, we used goodness of fit measures that take into account the number of parameters used in the models. That is, these measures--e.g., Akaike Information Criterion (AIC)--adjust the log-likelihood to favor more parsimonious models. Thus, our comparisons did not unduly or unfairly favor the more complete, 2-level models including SDU-level variables and cross-level interactions. Because each of six outcomes was examined in each of five modalities,

there were  $6 \times 5 = 30$  models in all. However, the introduction of SDU-level variables into the models for the correctional modality decimated the number of SDUs and cases due to missing data in the SDU-level variables. Consequently, further analyses on the correctional modality could not be performed.

### 3.0 RESULTS

#### 3.1 Main Effects

Exhibit 1 summarizes the variables with significant main effect coefficients ( $p < 0.10$ ) for each modality. The relatively liberal Type I error rate of 0.10 was selected for two basic reasons: 1) the tests were frequently underpowered, particularly for modalities with fewer SDUs (e.g., methadone), and 2) the analyses were exploratory in attempting to find SDU-level factors that might be important in explaining outcomes, rather than confirmatory tests of specific hypotheses. Note that a negative sign in the change outcome translates into increased improvement for each increasing level of the independent variable, and that a positive sign means reduced improvement in the severity scores. This direction follows from the way the change scores are constructed as post- minus pre-severity scores.

As shown, significant main effects of SDU-level variables were observed in three of the four modalities in which models could be run: methadone, non-methadone outpatient, and long-term residential. All six outcome domains were affected: drug, alcohol, criminal, employment, medical and psychiatric. In terms of factors associated with positive client change (i.e., increased reductions in severity scores), all but one were in the non-methadone outpatient modality.

Taken at face value, the data suggests that SDU-level main effects--as measured by these variables--were consistent across the six outcomes within a given modality, in that no variable had both positive and negative main effects within the same modality. However, they were inconsistent across modality, and effects representing reduced client improvement were about as likely to occur as effects representing increases. The one exception is having a designated case manager, which showed increases in improvement only, albeit in only one modality (non-methadone outpatient).

#### 3.2 Interaction Effects

Interactions occurred more frequently than main effects, suggesting that SDU-level characteristics do have a substantial impact on outcomes, but that impact is contingent on client-level characteristics. The pattern of significant interactions varies substantially by modality. As with main effects some interactions in the methadone and short-term residential modalities could not be tested due to lack of variance in SDU-factors.

Space does not permit a complete description of all the interactions, but some of the high points were as follows:

- The non-methadone outpatient modality had the most significant interaction effects (27), and was the only modality in which interactions were found for all 6 outcomes. The long-term residential modality had the next highest number (22), spanning 5 outcomes.
- Like the main effects, the signs of the interaction effects were consistent across outcomes within a given modality, but less consistent across modality. For example, the drug for which the client sought treatment frequently interacted with SDU-level factors, and did so in all four modalities. This strongly suggests that the client's particular drug problem mediates the effectiveness of SDU-level components, consistent with the literature on this topic (e.g., McLellan and Alterman, 1991; Thornton et al., 1998). However, the present study suggests that those mediational relationships may differ across modalities. For example, long-term residential clients presenting with a heroin problem showed greater improvements from patient matching in the drug and alcohol domain (relative to non-heroin users), while non-methadone outpatient clients with a heroin problem showed *less* improvement from patient matching in the drug, medical, and psychiatric domains. Treatment completion interacted positively with SDU-level factors in three of the four modalities.

In some cases, SDU-client interactions served to potentiate SDU main effects. For example, in the non-methadone outpatient modality, other-drug users benefited most from population tailoring in the medical domain, over and above the main effect of tailoring on medical severity. Similarly, clients in treatment due to legal pressure benefited most from population tailoring in the criminal domain, over and above the main effect of population tailoring on that outcome.

#### 3.3 Improvements in Model Fit

In the four modalities in which models could be tested, the introduction of SDU-level variables generally made little improvement in model fit, as assessed by the AIC statistic. In contrast, the addition of client by SDU interactions consistently improved model fit across all dependent. In the four modalities in which models could be tested, the introduction of SDU-level variables generally made little improvement in model fit, as assessed by the AIC statistic. The non-methadone outpatient modality showed greater improvements in absolute change, but was still very small in proportional change (less than 1 percent increase in pseudo- $R^2$  on all six outcomes). They were only modestly larger in the long-term residential modality, which (as noted above)

**EXHIBIT 1**  
**SIGNIFICANT SDU-LEVEL MAIN EFFECTS IN MIXED MODELS FOR EACH MODALITY AND**  
**OUTCOME VARIABLE**

<b>Dependent Variable Modality</b>	<b>Drug Severity</b>	<b>Alcohol Severity</b>	<b>Criminal Severity</b>	<b>Medical Severity</b>	<b>Employment Severity</b>	<b>Psychiatric Severity</b>
<b>Methadone</b>	Tailors to Pop. (+)					
<b>Non-methadone outpatient</b>	Case Management (-)	Case Management (-)  Training Specialists (+)	Case Management (-)  Training Specialists (+)  Tailors to Population (-)	Tailors to Population (-)  Average Individual Session Length (+)	Frequency of Individual Counseling (-)  Average Individual Session Length (+)	
<b>Short-term residential</b>						
<b>Long-term residential</b>	Training Specialists (-)  Frequency of Individual Counseling (+)				Tailors to Population (+)	

*Note: Coefficients that are significant at the 0.10 level are indicated, together with the sign (direction in parentheses) of the relationship. Minus signs indicate greater relationship.*

had the greatest number of significant effects. In contrast, the addition of client by SDU interactions consistently improved model fit across all dependent variables and modalities. Improvements in pseudo-R<sup>2</sup> ranged from a low of 1 percent (e.g., short-term residential, medical severity) to a high of 7 percent (e.g., long-term residential, drug and alcohol severity). In general, the largest improvements in fit were in the long-term residential modality, with the second largest in the methadone modality. While rules of thumb do not exist for interpreting how large an improvement in AIC is "large enough" to be important, we do note that the corresponding -2 log likelihood statistic showed a statistically significant improvement in most cases when the interaction terms were added and the AIC difference was modest (differences in the AICs are approximately equal to the log likelihood, which after multiplying by -2 can be used to test significant differences between models). For example, for changes

in drug severity for the non-methadone outpatient modality, models with and without interaction terms show a difference in -2 log likelihood of 8000 - 7813, or 187, which with 42 additional parameters is significant at p<.001.

#### 4.0 LIMITATIONS

The following limitations should be considered in interpreting these results:

- The large amount of missing data has repercussions for SDU-level analyses which must discard an entire SDU where values are missing for one or more of the SDU-level analysis variables (listwise deletion). Consequently, some of the modality-specific mixed models had limited degrees of freedom as we increased the number of (potential) explanatory variables. We compensated for this by making reasonable tradeoffs in limiting the number

of independent variables considered at the SDU level as well as the client level. Still (as noted above), the number of nonmissing cases remaining in the correctional modality was too small to support the modeling process. As a consequence, we cannot speak to the effects of SDU-level factors on clients in correctional treatment facilities. We also cannot speak to the effects of SDU-level factors with no variability in the methadone and short-term residential modalities.

- While the hierarchical models were specified correctly in Proc Mixed to the extent possible, one simplifying assumption was necessary for the procedure to successfully run. Specifically, it was necessary to assume that regression slopes of client-level predictors were similar across SDUs. The degree to which this assumption was violated, and its effect on the results if it was, are unknown.
- Some of the counterintuitive findings on SDU-level effects may in fact represent unmeasured selection bias. An example is the negative association between tailoring treatment to populations and outcomes in the methadone and long-term residential modalities. It seems unlikely that population tailoring would actually cause clients to do worse on outcomes; rather, the finding might be reflecting that SDUs with the most difficult clients tend to do more tailoring. By including the client-level variables first, outcome variance common to client-level factors and SDU-level factors was “credited” to client-level factors. This offered some measure of protection against the misinterpretation of selection bias as SDU-level factors effects, but only to the extent that the selection bias is captured by the client-level factors that were measured and included in the models.
- This analysis was based upon analysis of six outcomes in four treatment modalities, each in which 42 interactions were tested for significance along with main events. It should be expected that a portion of the significant results presented here could be due to chance alone.

## 5.0 IMPLICATIONS FOR TREATMENT RESEARCH, POLICY, AND PRACTICE

The results of our mixed model analyses show a range of significant SDU-level main effects and interactions between SDU variables and client-level variables. The results also highlight how these effects and interactions vary across modalities. These findings are possible with the multilevel modeling methodology that extends the multiple regression models previously conducted. The study therefore demonstrated some of the potential benefits in considering program effects in multilevel analyses of cross-site substance abuse

treatment evaluations, and in particular for NTIES data analyses. The following are areas for further research that could productively be investigated with the NTIES data:

- We view this report as a first attempt at applying multilevel random-effects models to the NTIES data, and it may well be that further refinements could yield more readily interpretable results. Given the limitations caused by missing data, however, it might be worth investigating the degree to which missing values on some SDU-level characteristics could be logically or statistically imputed from the response pattern on other, non-missing variables, or even recontacting the original SDU administrators who filled out the NBAR.
- In principle, a 3-level random effects analysis could be conducted that integrated the different modalities into a single model for each dependent variable. That is, clients and SDUs would still represent levels 1 and 2, respectively, and modality would represent level 3. This would permit a direct test of the effect of modality on outcomes, controlling for differences in both client case mix and SDU-level characteristics, as well as accounting for the intraclass correlations of clients within SDUs and SDUs within modality. It could also address interactions of modality with both SDU-level and client-level factors.

Though this study was primarily a methods demonstration, it nonetheless has substantive implications for policy and practice. In particular, the identification of SDU characteristics with positive effects and SDU by client interactions across different modalities and outcomes suggests ways in which the treatment community might refine programs along these dimensions. For example:

- The finding that positive main effects of SDU characteristics were primarily found in the non-methadone outpatient modality suggests that the likelihood of improving client outcomes through manipulating SDU-level factors (e.g., having a designated case manager and tailoring the program to the population) may be greatest in that modality.
- The finding that frequent, short sessions may be more beneficial than longer, less frequent sessions in outpatient treatment has implications for structuring outpatient programming.
- The finding that interactions occurred more frequently than main effects suggests that SDU-level effects are highly contingent on client-level characteristics and modality—confirming that in substance abuse treatment, the question of “what works” is more productively specified as “what works for whom, and in what setting?”

Other implications can be drawn from the findings highlighted above. However, as noted earlier, some of

the significant results presented here could be due to chance due to the large number of main effects and interactions examined. The optimal next step for policy and practice would be new prospective studies that rigorously examined some of the relationships suggested here.

## REFERENCES

- Cox, G.B.; Mejer, L.; Carr, D.I.; and Feng, S.A. (1993) Systems Alliance and Support (SAS): A program of intensive case management for public inebriates: Seattle. In : Connrad, K.J.; Hultman, C.I.; and Lyons, J.S. eds. *Treatment for the Chemically Dependent Homeless; Theory and Implementation in Fourteen American Projects*. New York : Hawthorne Press., Pp125-138.
- CSAT. (1994). Treatment Improvement Protocol (TIP) Series, 10. Assessment and Treatment of Cocaine-Abusing Methadone Maintained Patients. U.S. Department of Health and Human Services.
- CSAT. (1995). Treatment Improvement Protocol (TIP) Series, 20. Matching Treatment to Patient Needs in Opioid Substitution Therapy. U.S. Department of Health and Human Services.
- De Leon, G. Retention in drug-free therapeutic communities. (1991). In R.W. Pickens, C.G. Leukefeld, and C.R. Schuster (Eds.), *Improving Drug Abuse Treatment*. NIDA Research Monograph 106 (pp. 218-244). (NIH Publication No. 91-1754). Washington, DC: U.S. Government Printing Office.
- Hser, Y.I. (1995). Drug treatment counselor practices and effectiveness: an examination of literature and relevant issues in a multilevel framework. *Evaluation Review*, 19(4), 389-408.
- McCusker, J., Stoddard, A., Frost, R., & Zorn, M. (1996). Planned versus actual duration of drug abuse treatment: Reconciling observational and experimental evidence. *Journal of Nervous and Mental Disorders* 184(8), 482-489.
- McLellan AT; Alterman AI (1991). Patient treatment matching: a conceptual and methodological review with suggestions for future research. In *NIDA Research Monograph*, No. 106. Rockville, MD: National Institute on Drug Abuse.
- Mejta, C.L.; Bokos, P.L.; Mickenberg, J. H.; Maslar, E.M.; Hasson, A.L.; Gil, V; O'Keefe, Z; Martin, S.S.; Isenberg, H.; Inciardi, J. A. ; Lockwood, D.; Rapp, R.C.; Siegal, H.A.; Fisher, J.H.; and Wagner, J.U.H. (1994) Approaches to case management with substance abusing populations. In Lewis, J., ed. *Addictions Concepts and Strategies for Treatment*. Gatersburg, MD: Aspen.
- NORC. (1997). *The National Treatment Improvement Evaluation Study Final Report*. Prepared under contract by the the National Opinion Research Center (NORC) and Research Triangle Institute (RTI) for the Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.
- Orwin, R.G., Iachan, R., Ellis, B., Wolters, C. (1999a) *The National Treatment Improvement Evaluation Study: Multilevel Reanalysis Of Treatment Outcomes*. Prepared under contract by the Battelle Centers for Public Health Research and Evaluation for the Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.
- Orwin, R.G., Garrison-Mogren, R., Jacobs, M.L., & Sonnefeld, L.J. (1999b) Retention Of Homeless Clients In Substance Abuse Treatment: Findings From The NIAAA Cooperative Agreement Program. *Journal of Substance Abuse Treatment*, 17(1) 45-66.
- Orwin, R.G., Williams, V.F., & Ellis, B. (1999c) *The National Treatment Improvement Evaluation Study: Retention Analysis*. Prepared under contract by the Battelle Centers for Public Health Research and Evaluation for the Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.
- Orwin, R.G., Sonnefeld, L.J., Garrison-Mogren, R., & Smith, N.G. (1994) Pitfalls in evaluating the effectiveness of case management programs: Lessons from the NIAAA Community Demonstration Program. *Evaluation Review*, 18(2), 153-207.
- Polinsky ML; Hser YI; Grella CE (1998) Consideration of special populations in the drug treatment system of a large metropolitan area. *Journal of Behavioral Health Services Research*, 25(1):7-21
- Singer, J. D. (1999). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *The Journal of Educational and Behavioral Statistics*. (In press).
- Thornton CC; Gottheil E; Weinstein SP; Kerachsky RS (1998) Patient-treatment matching in substance abuse. Drug addiction severity. *Journal of Substance Abuse Treatment*, 15(6):505-511.
- Westermeyer J (1990) Treatment for psychoactive substance use disorder in special populations: issues in strategic planning. *Advances in Alcohol Substance Abuse*, 8(3-4):1-8.