DOCUMENT RESUME

ED 449 394	CG 029 842
AUTHOR	Koenig, Lane; Fields, Errol L.; Dall, Timothy M.; Ameen, Ansari Z.; Harwood, Henrick J.
TITLE	Using Case-Mix Adjustment Methods To Measure the Effectiveness of Substance Abuse Treatment: Three Examples Using Client Employment Outcomes.
INSTITUTION	Caliber Associates, Fairfax, VA.; National Evaluation Data Services, Rockville, MD.; Lewin Group, Inc., Washington, DC.
SPONS AGENCY	Substance Abuse and Mental Health Services Administration (DHHS/PHS), Rockville, MD. Center for Substance Abuse Treatment.
PUB DATE	2000-03-00
NOTE	46p.
CONTRACT	270-97-7016
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Adjustment (To Environment); Data Analysis; *Employment; Evaluation Methods; Models; *Multiple Regression Analysis; Outcomes of Treatment

ABSTRACT

This report demonstrates three applications of case-mix methods using regression analysis. The results are used to assess the relative effectiveness of substance abuse treatment providers. The report also examines the ability of providers to improve client employment outcomes, an outcome domain relatively unexamined in the assessment of provider effectiveness. This outcome is measured as the change between the number of days clients were paid for work in the 30 days prior to the intake interview and the 30 days before the follow-up interview. Consistent with previous research, the results confirm the need to use case-mix adjustment methods when assessing provider effectiveness. Although researchers may have long been aware of this finding, it is now crucial that those involved in the assessment of treatment providers recognize the importance of case-mix adjustment. Analyses accounting for differences in client characteristics reduce the risk of drawing inappropriate conclusions regarding the effectiveness of substance abuse treatment, thus limiting the possibility that incorrect treatment and treatment funding decisions are made. In addition, results show that estimates of provider rankings varied little across the three regression models when controlling for case mix, suggesting that provider rankings are not especially sensitive to choice of method. Implications for research, policy, and practice are discussed. (Contains 1 figure, 3 tables, and 27 references.) (MKA)

NEDS

NATIONAL EVALUATION DATA SERVICES

USING CASE-MIX ADJUSTMENT METHODS TO MEASURE THE EFFECTIVENESS OF SUBSTANCE ABUSE TREATMENT: THREE EXAMPLES USING CLIENT EMPLOYMENT OUTCOMES

March 2000

CSAT
Center for Substance
Abuse Treatment
SAMHSA

The LEWIN Group

NEDS

NATIONAL EVALUATION DATA SERVICES

USING CASE-MIX ADJUSTMENT METHODS TO MEASURE THE EFFECTIVENESS OF SUBSTANCE ABUSE TREATMENT: THREE EXAMPLES USING CLIENT EMPLOYMENT OUTCOMES

Principal Authors:

Lane Koenig, Ph.D.
Errol L. Fields, B.A.
Timothy M. Dall, M.S.
Ansari Z. Ameen, Ph.D.
Henrick J. Harwood
The Lewin Group

Series Editor:

Richard Finkbiner, Ph.D.
Caliber Associates

March 2000

CSAT
Center for Substance Abuse Treatment
SAMHSA

This document was supported by the Center for Substance Abuse Treatment, Department of Health and Human ...s, Caliber/NEDS Contract No. 270-97-7016.

TABLE OF CONTENTS

Page

FOREWORD i

ACKNOWLEDGMENTS ii

EXECUTIVE SUMMARY iii

I. INTRODUCTION 1

II. CASE-MIX ADJUSTMENT 3

III. STUDY DESIGN 6

IV. METHODS 8

1. MODELS AND OUTCOME MEASURES 8

Model 1: Ordinary Least Squares 8
Model 2: Logistic Regression 8
Model 3: Ordered Logistic Regression 9

2. MODEL SPECIFICATION 9

2.1 Client Demographics 10
2.2. Client Severity 10
2.3. Employment Counseling (Treatment Readiness) 11

3. RANKING PROVIDERS 11

V. DATA 13

1. DATA SOURCE: THE TRI EVALUATION DATABASE 13

2. INSTRUMENT: THE ADDICTION SEVERITY INDEX (ASI) 13

VI. RESULTS 14

1. SAMPLE DESCRIPTION 14

2. REGRESSION RESULTS 16

3. PROVIDER RANKINGS 18

3.1 Adjusted versus Unadjusted Rankings 18
3.2. Consistency in Rankings Across Models 20

TABLE OF CONTENTS (CONT.)

Page

VII. SUMMARY AND IMPLICATIONS FOR RESEARCH, POLICY, AND PRACTICE 21

1. SUMMARY 21
2. IMPLICATIONS FOR RESEARCH 22
3. IMPLICATIONS FOR POLICY 23
4. IMPLICATIONS FOR PRACTICE 23

REFERENCES 25

FOREWORD

The Center for Substance Abuse Treatment (CSAT) works to improve the lives of those affected by alcohol and other substance abuse, and, through treatment, to reduce the ill effects of substance abuse on individuals, families, communities, and society at large. Thus, one important mission of CSAT is to expand the knowledge about the availability of effective substance abuse treatment and recovery services. To aid in accomplishing that mission, CSAT has invested and continues to invest significant resources in the development and acquisition of high quality data about substance abuse treatment services, clients, and outcomes. Sound scientific analysis of this data provides evidence upon which to base answers to questions about what kinds of treatment are most effective for what groups of clients, and about which treatment approaches are cost-effective methods for curbing addiction and addiction-related behaviors.

In support of these efforts, the Program Evaluation Branch (PEB) of CSAT established the National Evaluation Data Services (NEDS) contract to provide a wide array of data management and scientific support services across various programmatic and evaluation activities and to mine existing data whose potential has not been fully explored. Essentially, NEDS is a pioneering effort of CSAT in that the Center previously had no mechanism established to pull together databases for broad analytic purposes or to house databases produced under a wide array of activities. One of the specific objectives of the NEDS project is to provide CSAT with flexible analytic capability to use existing data to address policy-relevant questions about substance abuse treatment. This report has been produced in pursuit of that objective.

Sharon Bishop
Project Director
National Evaluation Data Services

ACKNOWLEDGMENTS

The principal authors of this report were Lane Koenig, Ph.D., Errol L. Fields, Timothy M. Dall, M.S., Ansari Ameen, Ph.D., and Rick Harwood of The Lewin Group. We also wish to acknowledge our reliance upon the guidance and direction of Ron Smith, Ph.D., Program Evaluation Branch, the Government Project Officer for the NEDS contract. We also wish to thank Richard Finkbiner of Caliber Associates. Caliber Associates is the prime contractor for NEDS in partnership with Battelle Centers for Public Health Research and Evaluation (CPHRE); The Lewin Group; and National Opinion Research Corporation (NORC). Many individuals within the NEDS team contributed significantly to this report.

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

1. INTRODUCTION

The increasing emphasis on fiscal responsibility and accountability has led the Federal government, States, and managed care entities to increase efforts to identify cost-effective health care providers. Such efforts are evident in the field of substance abuse treatment where, increasingly, private and public payers are implementing initiatives to monitor the performance of providers. Faced with increased financial pressures to improve outcomes with fewer resources, providers are also recognizing the need to evaluate and monitor the relative effectiveness of their own treatment programs. In this report, we address some of the potential challenges of measuring the effectiveness of substance abuse treatment across providers and highlight the importance of controlling for differences in the characteristics of clients treated by each provider (i.e., provider case mix).

2. METHODS

We demonstrate three regression techniques (ordinary least squares, logistic, ordered logistic) using case-mix adjustment methods to evaluate the relative effectiveness of providers of outpatient treatment. A different construction of the same employment outcome measure was used in each model. In addition, all three models were estimated with and without controlling for differences in the characteristics of clients across providers (i.e., with and without case-mix adjustment). We used the results from each of the regression models to rank providers based on their estimated effectiveness at improving client outcomes. This approach allowed us to show: 1) the general applicability and importance of using case-mix adjustment methods, 2) several different types of approaches available to analysts, and 3) how the construction of the outcome measure and choice of statistical technique can affect estimates of provider effectiveness.

We used data collected by the Treatment Research Institute (TRI), a non-profit research institute working in collaboration with researchers at the University of Pennsylvania and the Veterans Administration, to compare provider effectiveness based on the change in the number of days clients were paid for work in the 30 days prior to treatment intake and the 30 days prior to the 6-month follow-up interview. The regression models were fitted using information for 1064 clients receiving outpatient treatment from 24 different providers.

The first model was estimated by ordinary least squares using the actual change in days paid as the dependent variable. We next estimated a logistic regression model and an ordered logistic regression model using categorical dependent variables. For the logistic model, the dependent variable indicates only whether or not there was an increase in the number of days a client was paid for work. For the ordered logistic model, we created a dependent variable that indicates if there was an increase, no change, or decrease in the number of days paid for work.

3. RESULTS

We estimated each of the three models with and without adjusting for differences in provider case mix. We then identified those providers who appeared to be statistically different, in terms of effectiveness, than the median-ranked provider, our proxy for the "average" provider. Provider rankings changed substantially when adjusted for case mix; the largest change in rank occurred in the ordinary least squares model where the 13th ranked provider (out of 17) in the unadjusted model climbed to the 1st position after adjusting for differences in case-mix. In the ordinary least squares and ordered logistic models, several providers were identified as outliers (i.e., either performing at a statistically significant level above or below the median-ranked provider) in the case-mix adjusted models, who were initially not identified as such in the unadjusted models.

When we controlled for differences in case mix across providers, four providers consistently ranked in the top four across all models, while three providers consistently ranked in the bottom three in the ordinary least squares and logistics models. Based on the ordinary least squares model, the top three ranked providers and the bottom ranked provider are statistically different from the median-ranked provider in terms of treatment effectiveness. In the logistic model, there were no outlying providers, while the top four providers in the ordered logistic model were found to be more effective than the median-ranked provider. Nevertheless, we observe a fair amount of consistency in the rankings across the three case-mix adjusted regression models. The rankings of 7 providers (41 percent) do not vary by more than one place across the three models, and only 4 providers (24 percent) vary in rank by 4 or more places.

4. DISCUSSION AND IMPLICATIONS

This analysis contributes to building a case for helping treatment systems and providers to be accountable for their performance. Both consumers and those paying for treatment (public agencies and private insurance) want to know that more effective and less effective providers can be identified, in order to learn from the former and improve the latter. The findings validate the concern of providers that different clients have different expected or predicted outcomes, and providers with more difficult clients need to be viewed differently than those with less severe clients. Providers should be aware that performance measurement efforts are gaining momentum and they need to engage in the process by which these measurement systems are being developed and implemented in order to inform and shape the process and the system. While researchers have developed a variety of ways in which performance analysis can use case-mix adjustment, the basic approach appears to be very appropriate to the assessment of substance abuse providers.

The variation in provider rankings between the case-mix adjusted and unadjusted models confirms the need to use case-mix adjustment methods. Without adjusting for case mix, an evaluator may incorrectly conclude that a particular provider is more or less effective at improving a particular outcome than other providers. Despite modeling the changes in client employment status using three alternative specifications of the dependent variable and using different estimation techniques, provider rankings varied little across the three models. It should give some comfort to evaluators and providers that the rankings of providers in our example remained fairly stable across models. Although rank orderings differed very little, our models did identify different sets of providers who were either "statistically" more or less effective than the median-ranked provider, with the logistic model identifying the smallest number of providers whose effectiveness differed significantly from the median-ranked provider. In this respect, the ordinary least squares and ordered logistic models allowed for greater differentiation between providers and, thus, appear to be superior to the logistic model for evaluating provider effectiveness.

I. INTRODUCTION

I. Introduction

The increasing emphasis on fiscal responsibility and accountability has led the Federal government, States, and managed care entities to increase efforts to identify cost-effective health care providers. Such initiatives are spreading beyond general health care into the fields of behavioral health, such as substance abuse treatment, where private and public payers are increasingly adopting systems for evaluating the effectiveness of providers. Faced with increased financial pressures to improve outcomes with fewer resources, providers are also recognizing the need to evaluate and monitor the relative effectiveness of their own treatment programs.

Systems for assessing provider effectiveness can serve several functions, including benchmarking, evaluating the impact of program-level changes on client outcomes, identifying exceptional providers to uncover best practices, and identifying candidates for continuous quality improvement processes.[1] Despite the need for information on methods to accurately measure treatment effectiveness, there exists relatively little published literature on how to assess the performance of substance abuse treatment providers. In practice, evaluators attempting to measure provider effectiveness face several major challenges, including identifying fair, appropriate and efficient measures of performance, and identifying and employing appropriate techniques to compare providers.

This report addresses some of these challenges by illustrating three regression techniques to assess the relative effectiveness of outpatient substance abuse treatment providers. Using measures of employment as our outcome variable, we estimated each model with and without adjusting for differences in the characteristics of clients treated by each provider (i.e., with and without adjusting for provider case mix) and then ranked providers according to their estimated treatment effectiveness.[2] While, for purposes of this report, we measured effectiveness using an employment outcome measure, it is important to recognize that the approaches demonstrated in this paper can be applied to other outcome measures. However, by estimating three regression models using different constructions of a single client outcome measure, we were able to compare the results from each model to demonstrate the effects of using case-mix adjustment methods and the range of approaches available to evaluators. Moreover, we hope to show how

[1] For a general discussion of the importance of evaluating provider effectiveness using case-mix adjustment methods, see Harwood et al. (1997).

[2] While case-mix adjustment methods are most commonly discussed in the literature for comparing the relative effectiveness of providers or ranking provider performance, it is important to recognize that any evaluation of treatment effectiveness that is based on outcomes across different providers must account for differences in case mix.

the construction of the outcome measure affects the estimates of provider effectiveness. In particular, we are interested in addressing the following questions:

1. How does using case-mix adjustment methods affect estimates of provider treatment effectiveness when examining employment outcomes?

2. Do our estimates of provider effectiveness depend on how we measure client employment outcomes?

3. Do the regression models identify different sets of outliers (i.e., those providers more or less effective than the "average" provider)? If so, in what ways do the set of outliers change?

II. Case-Mix Adjustment

II. Case-Mix Adjustment

Case mix refers to the characteristics of cases served by a health service provider, where some clients are at greater "risk" of having less successful treatment outcomes than other clients. In drug abuse treatment, level of risk is commonly associated with addiction severity, but other factors may also be important, such as client demographics, socioeconomic status, and medical and social functioning. Substance abuse treatment providers often serve clients who differ dramatically along these risk factors, frequently specializing in the treatment of certain client populations.

Since differences in the types of clients treated across providers result in different expected treatment outcomes, assessing the relative treatment effectiveness across providers without adjusting for client differences may result in spurious and misleading findings. In fact, any attempt to accurately measure the cost-effectiveness of substance abuse treatment, the cost-effectiveness of different treatment services, or the relative effectiveness of different treatment settings using client outcomes across different providers needs to account for differences in the clients treated to ensure the validity of the findings.

Case-mix adjustment (CMA) is a tool that enhances the accuracy and quality of assessments of provider or treatment effectiveness by controlling for those client-level factors that *affect* client outcomes but that are beyond the control of providers.[3] There exist a number of benefits to using CMA, including:

- Increasing the validity of assessments of client functioning;
- Increasing the validity of assessments and comparisons of provider effectiveness; and
- Establishing realistic performance benchmarks for provider effectiveness that take into account client severity and functioning.

A review of the medical and substance abuse literature regarding the use of CMA methodology reveals that CMA has been used primarily within the context of analyzing hospital performance. Within the context of the hospital literature, for example, Hannan et al. (1991, 1994, and 1995), and Luft and Romano (1993) ranked hospitals using case-mix adjusted mortality rates of clients who received a coronary artery bypass graft (CABG). Smith, McFall, and Pine (1993) similarly used CMA to identify differences across states in mortality rates of hospitalized Medicare beneficiaries. To a much lesser extent, the CMA literature has focused on

[3] When considering client-level controls in CMA analysis, it is important to only identify those factors that can be expected to directly influence outcomes independent of provider action.

substance abuse treatment and providers. This small but growing body of literature on CMA methods within the substance abuse arena has focused on:

- Development and refinement of outcomes that can be attributed to substance abuse treatment;
- Development and refinement of the instruments needed to collect outcomes-related information, such as the Addiction Severity Index (ASI);
- Identification and application of appropriate methods for making provider comparisons (e.g., Phillips et al., 1995; Dall et al., 1999); and
- Identification of parsimonious models upon which case-mix adjustment models can be based (e.g., Ameen et al., 1999).

In the substance abuse literature, some researchers have focused on ranking providers based on client outcomes after adjusting for client risk factors at intake (e.g. Phillips et al., 1995), while others focus more attention on linking client outcomes to the nature and amounts of treatment services provided (e.g. McLellan et al., 1993). Using two inpatient and two outpatient private treatment providers, McLellan et al. (1993) addressed the question of whether some providers are more effective than others. The authors controlled for client severity at treatment intake in six areas including medical status, employment and self support, alcohol and drug use, legal status, family and social relationships, and psychiatric symptoms, and found consist differences across providers in client social functioning and substance abuse six months following treatment intake.

Phillips et al. (1995) compared the performance of 18 methadone providers. Their focus was on ranking providers rather than measuring the nature and amounts of treatment services provided during treatment. The authors used logistic regression to predict six client outcomes—heroin use, cocaine use, employment, arrests, depression, and retention at three months into treatment. Risk factors included in the model were age, gender, race, education, mental health, drug use history, drug and mental health treatment history, and employment and arrest history. Using the results from a series of logistic regressions, providers were ranked on the basis of their estimated performance. Provider rankings for each outcome were then averaged to derive a measure of overall performance. Results generally confirmed that, for each domain, client severity at intake tends to be a significant predictor of outcomes three months after treatment intake. Although the data used in the analysis are 15 years old and outcomes are not measured post-discharge, the study's findings underscore the importance of controlling for difference in client severity across providers (i.e., performing case-mix adjustments) when making provider comparisons.

Phibbs et al. (1997) employed case-mix adjustment methodologies to compare provider effectiveness on the basis of readmission rates across 116 Veterans Affairs Medical Centers between 1987 and 1992. Because direct measures of client outcomes, such as reductions in substance abuse, relapse rates, employment and legal status, etc., were unavailable, Phibbs et al. used readmission rates as their outcome measure. The authors used logistic regression to predict readmission to treatment with controls for risk factors grouped into demographic characteristics, psychiatric and medical comorbidities, and type of substance abuse. Providers were ranked on the basis of the ratio of each provider's actual to predicted readmission rates. Results confirmed that CMA led to a significant re-ordering of provider rankings.

A recent analysis by Ameen et al. (1999) used data collected by the Treatment Research Institute using the Addiction Severity Index (ASI) to rank providers based on changes in their clients' drug use days from intake to follow-up. The authors controlled for variations across providers in client demographics (gender, age, race, education, and marital status), client functioning in the seven domains of functioning measured by the ASI (drug use, alcohol use, employment, family, medical, legal and psychiatric), and a proxy for clients' readiness for treatment (client-reported importance of treatment for drug problems). The authors found a significant re-ordering of provider effectiveness rankings when they compared the non-case-mix-adjusted rankings with the case-mix adjusted provider rankings. These results further underscore the importance of adjusting for client characteristics, including presenting problems and degree of severity, to increase the validity of estimates of provider effectiveness.

III. Study Design

III. Study Design

This report demonstrates some of the case-mix adjustment methods available to assess the relative effectiveness of substance abuse treatment providers. We used three constructions of the same employment outcome measure to compare the effectiveness of providers of outpatient services. Each outcome was constructed so that a different regression technique would be appropriate for each employment outcome measure. In addition, each model was estimated with and without controlling for differences in the characteristics of clients across providers (i.e., with and without case-mix adjustment). This approach allowed us to illustrate 1) the general applicability and importance of using case-mix adjustment methods when measuring the effectiveness of substance abuse treatment, 2) some of the different types of case-mix adjustment approaches available to analysts, and 3) how the construction of the outcome measure and the choice of statistical method can affect the estimates of provider effectiveness.

We used three regression techniques to rank providers along a single dimension: the ability to improve the employability of clients. We chose employment as our performance measure because it captures improvements in clients' personal health and social functioning, a logical outcome of accessing substance abuse treatment.[4] Improvements in employment are important to stakeholders and directly benefit both clients and society, through increases in income and income taxes and reductions in welfare expenditures, crime, and criminal justice expenditures. Moreover, we believe that employment outcome measures meet the four criteria for evaluating outcome measures identified in Burnam (1996): it is well-suited to populations and purposes, it has good psychometric properties, the burden and cost to collect the information is minimal, and it is clearly interpretable. Despite its societal importance and its suitability as an outcome measure, the use of employment outcome measures to assess the performance of substance abuse treatment providers has been limited.[5]

Within the regression analysis framework, there are many different techniques available to analysts.[6] The most appropriate technique to use depends on, among other factors, the form

---

[4] McLellan et al. (1992) identify three major outcome domains that are relevant to the rehabilitative goals of patients and the public: 1) sustained reduction in drug and alcohol use, 2) sustained improvements in personal health and social function, and 3) sustained reductions in threats to public health and safety.

[5] It is important to recognize that evaluating employment outcomes is inappropriate for certain groups of clients, such as children and individuals institutionalized.

[6] A review of the substance abuse case-mix adjustment literature, however, finds that only a limited range of techniques has been used to date. The principal methods used in the substance abuse literature have been the ordinary least squares regression and the logistic regression (see for example, Ameen et al., 1999; Dall et al., 1999: Phibbs et al., 1997; and Phillips et al., 1995).

and distribution of the dependent variable in the regression equation (see Greene, 1997). As we demonstrate below, researchers can often characterize outcomes in several ways, with each characterization resulting in a different distribution of the dependent variable and, therefore, requiring a different regression technique. In practice, numerous considerations may influence how researchers specify the dependent variable, such as the ability to defend the model to key stakeholders, the policy relevance of the outcome, the model's interpretability, the availability of computing resources, and the analyst's familiarity with certain models.

Our outcome measure is based on client responses to an interview question that asks the number of days clients were paid for work in the past 30 days. The question is asked at intake and again during the follow-up interview. Using information on the change in number of days paid for work between the intake and follow-up reference periods, we created three different employment outcome measures and estimated the most appropriate model for each one. In each model, we adjusted for differences in case mix across providers by including in the regression equations explanatory variables for client characteristics and presenting conditions. We then compared the provider rankings obtained from each of the models.

While there are, of course, several other employment outcome variables that one could use, such as a variable indicating whether or not a client was employed, our choice of employment outcome measure was based on two considerations. First, a variable equal to the change in days paid for work has features similar to other outcome variables used in substance abuse research; that is, it is bounded (between –30 and 30), discrete, and observations tend to be concentrated around a single or small number of values.[7] Second, it allows flexibility in specifying the dependent variable in the regression equations as demonstrated below.

[7] Other examples of outcome measures with similar characteristics are the number of days a client used drugs in the last 30 days or the number of times a client was arrested.

IV. Methods

IV. METHODS

In this section, we discuss our outcome measures, methodology, model specification, method for identifying providers significantly different from the "average," and our approach to ranking providers. We first consider the different regression models and various constructions of the outcome measures.

1. MODELS AND OUTCOME MEASURES

We estimated an ordinary least squares model using a semi-continuous (i.e., bounded between –30 and 30) variable based on the change in days paid for work between the 30 days prior to the follow-up interview and the 30 days prior to the intake interview. We estimated a logistic regression model and an ordered logistic regression model using categorical dependent variables based on the change in days paid.[8] The dependent variables and associated models are discussed in more detail below.

Model 1: Ordinary Least Squares

The first choice of a dependent variable was the change in days paid for work. This variable was calculated by subtracting the number of days paid for work during the 30 days prior to treatment intake from the number of days paid for work during the 30 days prior to the follow-up interview (which occurred approximately 6 months following treatment intake). The transformed variable has a distribution that is semi-continuous with end points at –30 and 30 and approximates a normal distribution centered on zero. For this dependent variable the ordinary least squares model is the most appropriate. One advantage of using this measure of employment change is that it incorporates all the available information on the dependent variable, whereas the logistic and ordered logistic regression models collapse the data into categories.

Model 2: Logistic Regression

An alternative to using the actual change in the number of days paid is to create outcome categories that summarize this information. For Model 2, we examined the change in days paid for each client and recorded a client as "improved" if the number of days increased or "no

[8] While we chose to work with the change in days paid, an alternative specification would be to use information on the number of days paid in the follow-up period as the dependent variable and information on the number of days paid prior to intake as a control variable.

improvement" if the number of days did not change or decreased. For a dichotomous dependent variable such as this, the most frequently employed model is the logistic regression model.[9]

Model 3: Ordered Logistic Regression

For Model 3, we created a dependent variable with three categories. As with the outcome measure described for Model 2, we created one category we label "improved," which includes clients for whom the number of days paid increased in the post-treatment period. We then divided the category "no improvement" into two: "no change" and "worsened." Using these three categories for the dependent variable, we estimated an ordered logistic regression model to rank providers.

Ordered logistic regression models are cited less frequently in the substance abuse treatment literature than logistic regression models, but are nevertheless very useful for analyzing models where the dependent variable may assume more than two categories and where there is a natural ordering to the categories. Ordered-categorical data are common in survey data. For example, a survey of clients receiving substance abuse treatment may report the health status of clients as poor, fair, good, or excellent, or the number of crimes committed by a client may be recorded as zero, 1-5, 6-10, or more than ten.

One way of dealing with categorical data of this type is to collapse the categories into two and estimate the transformed data using the standard logistic regression model. However, this is not an efficient use of the available data. In this sense, the ordered logistic regression is an improvement over the standard logistic regression model since it uses more information about the change in days paid for work.

2. MODEL SPECIFICATION

The three models contain the same set of explanatory variables. These variables were derived from a review of the literature on the determinants on employment outcomes and the conceptual framework used by Ameen et al. (1999) for identifying the key factors likely to affect client outcomes. The major clusters of factors include client demographics, client severity at

[9] The probit model is also designed to handle a dichotomous dependent variable, but has been less widely used in the substance abuse literature. The probit model assumes that the model's error term is normally distributed. Because the normal and logistic distributions are similar, the models are unlikely to produce very different results (Maddala, 1983).

intake in several domains (substance abuse, social functioning, etc.), and a measure of the importance of counseling for employment problems.[10]

2.1 Client Demographics

The models used in this analysis include basic demographic characteristics, including gender, age, marital status, education, race, and ethnicity. We included these demographic variables in the model because labor force participation differs among individuals with different demographic characteristics. For example, married women, particularly women with children, are more likely not to work outside the home than men or single women. In addition, different groups may exhibit different abilities to find work as well as differences in the desire to obtain full-time or part-time work.

Age and education are both continuous variables with values measured in years. Marital status is a dichotomous variable where 1= married and 0 = not married (including separated, divorced, widowed, and never married). Gender is a dichotomous variable where 1 = female and 0 = male. Race is also a dichotomous variable where 1= black (not of Hispanic origin) and 0= not black. Ethnicity is described by a dichotomous variable where 1= Hispanic and 0 = not Hispanic.

2.2. Client Severity

The ASI composite scores at intake were included as independent variables in the analysis. The composite scores were designed as general status measures of each problem area. Their inclusion allows for the possibility that, regardless of cause, the initial level of functioning in the medical, alcohol, drug, legal, family/social, and psychiatric areas may affect employment outcomes. However, the models also included days worked in the past 30 (at intake) to account for the statistical artifact that those individuals with little paid work at intake can experience the greatest increase in the number of days paid for work. Therefore, the ASI employment composite, which is computed using the days worked in past 30, was omitted to avoid the problem of multicollinearity. Each composite score is the sum of answers to several questions within the six ASI domains. The individual items are not weighted since there is no theoretical, empirical, or clinical justification for establishing a weighting scheme. Mathematical adjustments account for the different response ranges of the questions and the number of items

[10] Ideally, the model should also include information about the local job market and information about a client's training and skills, since work opportunities will be different across occupations. We were limited in our choice of explanatory variables by the availability of the data.

in the composite. The composite scores have been recommended for use in studies assessing change (Alterman et al., 1994) with previous research demonstrating their usefulness (Campbell, 1997).

2.3. Employment Counseling (Treatment Readiness)

Recent theoretical and empirical developments with regard to the impact of motivation on outcomes suggest the need to include a measure of the importance of treatment/counseling when predicting outcomes. More specifically, the Transtheoretical Model (Prochaska & DiClemente, 1986, 1992; Velicer et al., 1995) has gained in ascendancy in recent years and suggests that clients present for treatment at different stages of readiness, which in turn affects outcomes. The Transtheoretical Model also shows promise in the field of substance abuse: evidence suggests that a client's initial motivation and readiness are related to short-term retention in therapeutic communities (DeLeon, Melnick & Kressel, 1997) and that a client's willingness to enter treatment positively influences drug use outcomes, treatment tenure, and housing outcomes in a sample of homeless adults (Erickson et al., 1995).

Although this theoretical model is more directly applicable to readiness for substance abuse treatment, the literature suggests an application to readiness for change regarding additional outcomes of treatment such as employment. Individuals who are motivated to improve their employment status may be more likely to attain positive employment outcomes than individuals who lack that motivation. In this analysis, the question on the drug/alcohol use section of the ASI that asks "how important to you now is counseling for employment problems" was used as a proxy for clients' readiness for treatment. This variable is constructed such that 0 = not at all, 1= slightly or moderately and 2 = considerably or extremely.

3. RANKING PROVIDERS

There are two approaches used to rank providers in the available case-mix adjustment literature. The most widely used approach is to aggregate predicted client outcomes and actual outcomes by provider and then compare actual to predicted outcomes to measure provider effectiveness. Alternatively, one can estimate a regression model that contains provider indicator variables (or "dummy" variables) to capture unexplained variation in the dependent variable that varies systematically by provider.[11] We used the latter approach for the reasons discussed in Dall

[11] A detailed discussion and comparison of the differing approaches is available in Dall et al. (1999).

et al. (1999)—namely, to control for omitted variable bias and to distinguish between within-provider and cross-provider variation in client outcomes.

For each of our models, we regressed the measure of employment change against a series of N-1 provider indicator variables and a set of client-level variables to account for differences in case mix across providers. When using indicator (also known as "dummy") variables, one group is designated as the reference group and the coefficient on each provider dummy variable indicates whether the client outcomes of that provider are expected to differ, on average, from client outcomes of the reference group after adjusting for differences in case mix. We used the median-ranked provider as the reference group, and omitted the provider dummy variable for this provider in the regression models.[12] Thus, a provider dummy variable that has a positive coefficient indicates that the provider was more effective at improving the employment outcomes of clients relative to the median-ranked provider. The reference provider may be chosen because it fits some measure of "average" effectiveness. However, our selection of the median-ranked provider as our reference provider is in no way meant to imply that the median-ranked provider provides "good" or "adequate" care and those ranked below it provide "poor" or "inadequate."

There are, of course, other possible criteria that may be used to identify the reference provider. For example, stakeholders may use the most costly treatment provider to see if less costly providers render more or less effective treatment. Individual providers may identify an exemplary provider against whom they might compare themselves. Over time, they may use the results of their CMA analysis to track changes in the effectiveness of their treatment and identify successful treatment strategies. The criteria for identifying the reference provider depends on the purpose of the analysis.

[12] This required us to first identify the median-ranked provider by running each regression once while excluding the indicator variable for an arbitrarily chosen provider and then ranking the providers to identify the median-ranked one. Instead of using the median-ranked provider, we could have chosen to compare providers to the overall average. This is easily done in the OLS model by including a dummy variable for all providers and restricting the parameters on the provider dummy variables to sum to zero. The resulting parameter on a provider dummy variable represents the effectiveness of that provider relative to the overall average. This simple approach, however, does not apply to the logistic and ordered logistic models.

V. Data

V. Data

1. DATA SOURCE: THE TRI EVALUATION DATABASE

Data used in this analysis were collected by the Treatment Research Institute (TRI), a non-profit research institute working in collaboration with researchers at the University of Pennsylvania and the Veterans Administration. The TRI evaluations follow a random sample of approximately 75-100 clients per program (usually consecutive admissions) using an intent-to-treat design. Under the intent-to-treat design a random sample of clients is selected at admission and fully assessed throughout treatment and the follow-up period. Data needed for evaluations are collected at intake and again at 6 months post-admission using the ASI. Well-designed follow-up procedures result in an average of 84 percent of all clients in TRI studies being successfully contacted at follow-up. Methods for insuring validity of client responses include urine and breathe samples from a random sample of twenty percent of subjects.

2. INSTRUMENT: THE ADDICTION SEVERITY INDEX (ASI)

Client data were collected using the ASI, a standardized clinical research interview that assesses problem severity in six domains commonly affected among substance abuse and mental health clients. The ASI's design makes it conceptually well suited for the purpose of ascertaining changes in client status from intake to discharge and follow-up. The instrument was originally developed and introduced in 1980 to evaluate treatment outcomes across different providers. The questions were designed to cover a broad range of problems that should be affected by substance abuse treatment (drug use/alcohol use, medical, employment, legal, family, and psychiatric) and to be amenable to repeat administrations. Multiple examinations of the ASI severity ratings and composite scores have produced evidence of its concurrent reliability and validity across subgroups of clients (McLellan et al., 1992). The updated fifth edition of the ASI was designed to keep pace with the dynamic nature of drug use as well as developments by substance abuse researchers (McLellan et al., 1992). Given their demonstrated influence on substance abuse patterns, the fifth edition of the ASI includes items that measure family influences, abuse relationships, levels of social support and psychiatric disorders.

VI. RESULTS

VI. Results

1. SAMPLE DESCRIPTION

We conducted the analysis using data on 1064 clients, distributed across 24 providers, who received treatment in an outpatient setting. For this illustrative analysis, we chose to analyze the effectiveness of providers who served outpatient clients rather than providers who treated clients in an inpatient or methadone treatment facility for two reasons. First, our sample was much larger for the outpatient modality, both in terms of clients and providers, than for either the inpatient or methadone modalities. Second, clients who receive treatment in an outpatient setting tended to have less severe substance abuse problems, and thus the goal of improving clients' employment situation is presumably more realistic for these clients.

The initial sample included information on 1440 clients who completed both an intake and follow-up interview. Clients in a controlled environment (i.e., jail or inpatient setting for drug, medical, or psychiatric treatment) for more than 14 of the 30 days prior to *intake* were excluded from our analysis. In addition, we also excluded clients who were in a controlled environment, other than jail or an inpatient setting for drug treatment, for more than 14 of the 30 days prior to the *follow-up* interview. These exclusions (n=195) were made to ensure the consistency of our comparisons across providers. However, clients who were in jail or alcohol/drug treatment prior to the follow-up interview and not in a controlled environment for more than 14 days in the month prior to intake were *included* in the analysis. The limited number of days worked by these clients prior to the follow-up interview due to incarceration or inpatient substance abuse treatment are legitimate, poor outcomes.[13] Client observations with missing values for any of the independent or dependent variables also were omitted from the analysis sample (n=181). In addition, for purposes of this illustrative analysis, eight providers with fewer than 15 clients each in the analysis sample (after all exclusions) were combined as a single provider (Provider F). We did this because the number of clients from each of these provider was insufficient to ensure accurate and stable measures of provider effectiveness. Thus, our analysis covers 17 "providers," 16 providers with 15 or more clients plus the composite

[13] We are grateful to an anonymous reviewer who pointed out that the exclusion of clients who were incarcerated during the follow-up period is not necessary, since such an outcome can be viewed as a negative result. In general, the decision on whether or not to exclude clients who were incarcerated during either the pre- or post-treatment periods depends on the outcome measured and the purpose of the study. For example, while it may make sense to include clients who were jailed following treatment in an analysis of employment outcomes, it may be inappropriate to include them in analysis of criminal activity. This is because the analysis will incorrectly consider this a positive outcome since the number of crimes committed by the client can not increase.

provider, F, which includes the clients for the eight remaining providers with less than 15 clients.[14]

Of the 1064 individuals in the final analysis sample, the average number of days employed of the 30 days prior to intake was 6.4 days, while the average number of days employed of the 30 days prior to follow-up was 8.5 days. On average, clients had completed 11.6 years of education (Exhibit VI-1). The majority of individuals believed that counseling for employment problems was not important (56%); a small percentage (10%) believed counseling was slightly or moderately important while a significant minority (33%) believed counseling was extremely or considerably important. Demographically, this population was predominately male (63%), unmarried (81%), African American (65%), and non-Hispanic (94%) with an average age of 36 years.

EXHIBIT VI-1
EXPLANATORY VARIABLES, MEANS AND STANDARD DEVIATIONS
(SAMPLE SIZE = 1064)

Variable	Mean	Std. Deviation
Gender	0.37	0.48
Age	35.59	8.55
Marital Status	0.19	0.39
African-American	0.65	0.48
Hispanic	0.06	0.24
Years of Education	11.56	1.98
Drug Composite Score	0.13	0.13
Family Composite Score	0.21	0.22
Legal Composite Score	0.06	0.14
Medical Composite Score	0.21	0.33
Alcohol Composite Score	0.29	0.29
Psychiatric Composite Score	0.24	0.25
Number of Days Worked Prior to Intake	6.41	9.46
Importance of Counseling for Employment Problems	0.77	0.92

Data source: TRI Evaluation Database, information collected using the ASI

[14] Alternatively, we could have eliminated these observations. However, we chose to include these observations because they provide additional information on which to base our case-mix adjustments.

2. REGRESSION RESULTS

Despite applying three different regression techniques, each with a different measure of employment, the estimates of the relationship between employment status and client characteristics are quite stable (see Exhibit VI-2). Across all models, our results show that clients with higher levels of education consistently show more improvements in employment status than clients with less education. This finding is statistically significant at the 0.05 level in all three models, although the magnitude of the effect is small. For example, in the ordinary least squares model two additional years of schooling are associated with an improvement of only one additional day of paid work.

Women show more improvement than men do in their employment status following treatment. This finding is statistically significant at the 0.05 level in all three models. The ordinary least squares model results show, on average, the improvement in days worked is four days greater for women than for men. These findings differ from those of Wright and Devine (1995) and Lapham et al. (1995) who find that women are less likely than men to have improved employment outcomes following treatment.

Clients with less severe medical problems (indicated by a higher medical composite score) show less improvement in employment status than clients with less severe medical problems. This finding is statistically significant at the 0.05 level in all three models. However, the magnitude of the effect is difficult to ascertain because the explanatory variable is a composite of clients' responses to multiple questions regarding medical problems.

Clients who report at treatment intake that counseling for employment problems is important or very important show more improvement than clients who report that the counseling is not important or only moderately important. This finding is statistically significant at the 0.10 percent level in the ordinary least squares model, and significant at the 0.05 level in both the logistic regression and ordered logistic regression models. These results lend further support to the Transtheoretical Model.

Age is negatively correlated with improvements in employment status in all three models; however the findings are not significant and the magnitude of the effect in each of the models is modest. Married clients show more improvement in employment status than unmarried clients. This finding is significant at the 0.05 level for the ordinary least squares model. African

EXHIBIT VI-2 ALTERNATIVE CASE MIX-ADJUSTED MODELS OF PROVIDER EFFECTS ON EMPLOYMENT			
	Model 1 OLS	Model 2 Logistic	Model 3 Ordered Logistic
Intercept	4.66** (2.24)	-1.00 (0.63)	-0.85 (0.64)
Intercept2	--	--	2.02** (0.64)
Days Paid Prior to Intake	-0.63** (0.04)	-0.05** (0.01)	-0.13** (0.01)
Importance of Counseling for Employment Problems	0.55* (0.31)	0.23** (0.08)	0.15** (0.07)
Male (Male-1, Female-0)	-3.89** (0.64)	-1.21** (0.19)	-0.79** (0.15)
Age	-0.05 (0.03)	-0.02 (0.01)	-0.01 (0.01)
Married (Married-1, Other-0)	1.49** (0.72)	0.19 (0.19)	0.27 (0.17)
African American (AA-1, Other-0)	-0.76 (0.74)	-.035* (0.20)	-0.35** (0.18)
Hispanic (Hispanic-1, Other-0)	-0.24 (1.55)	-0.53 (0.43)	-0.49 (0.37)
Years of Education	0.51** (0.14)	0.15** (0.04)	0.12** (0.03)
Drug Composite Score	-3.16 (2.42)	-0.55 (0.68)	-0.85 (0.58)
Family Composite Score	-0.07 (1.29)	0.15 (0.36)	0.04 (0.31)
Legal Composite Score	0.53 (1.90)	0.07 (0.50)	-0.24 (0.45)
Medical Composite Score	-3.41** (0.85)	-0.89** (0.25)	-0.71** (0.20)
Alcohol Composite Score	-1.19 (1.08)	-0.08 (0.30)	-0.08 (0.26)
Psychiatric Composite Score	0.19 (1.27)	0.30 (0.35)	0.28 (0.30)
Adjusted R^2	0.28	--	--
-2 Log L	--	139.53** (DF=30)	241.88** (DF=30)
Sample Size	1064	1064	1064

Notes: ** (*) indicates statistical significance at the .05 (.10) significance level. The dependent variable in Model 1 is the change in number of days worked between the 30 days prior to the follow-up interview and 30 days prior to the intake interview. For Model 2, we used a variable indicating whether or not the number of days increased. In Model 3, the dependent variable indicates if the number of days worked increased, decreased, or stayed the same.

Data source: TRI Evaluation Database, information collected using the ASI, analysis by The Lewin Group.

least squares and ordered logistic models, several providers, who were initially not identified as outlying providers (i.e., either performing at a statistically significant level above or below the median-ranked provider), were identified as outliers after adjusting for client demographics and severity of presenting symptoms. For example, Provider D although not statistically different from the medium-ranked provider before case-mix adjustment in either the OLS or ordered logistic models, is estimated to be above the median after adjusting for provider case-mix. Similarly, Providers C and O, while initially identified as outliers in the logistic models, are no longer significant after case-mix. Exhibit VI-4 illustrates the re-ordering of provider rankings that occurs after adjusting for provider case mix for the ordinary least squares model.

EXHIBIT VI-3
PROVIDER RANKINGS UNDER THREE DIFFERENT MODELS

Provider	Number of Clients	OLS		LOGISTIC		ORDERED LOGISTIC	
		Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
A	28	13	1**	3	1	2	1**
B	20	17	2*	12	4	13	3**
C	16	9	4	1*	2	1	2**
D	40	14	3**	5	3	12	4**
E	51	12	5	6	6	10	5
F	60	15	6	11	7	16	6
G	44	1	7	9	9	4	9
H	171	4	10	7	12	5	13
I	101	2	11	8	10	3	8
J	17	16	18	2	5	7	7
K	132	3	9	14	14	6	10
L	115	5	13	15	13	11	12
M	61	8	12	4	8	9	11
N	73	6	14	13	15	8	15
O	51	10	16	17**	16	15	14
P	31	11	15	10	11	17	16
Q	55	7	17*	16	17	14	17

Note: The omitted provider in the adjusted ordinary least squares, logistic and ordered logistic regression models were Providers K, G, and G, respectively. The omitted provider for the unadjusted ordinary least squares, logistic and order logistic regression models were Providers C, G, and M respectively. Thus, their implicit parameter values are zero, and their rankings were based on this value. **(*) Indicates a provider's effectiveness is significantly different from the median-ranked provider's effectiveness at the .05 (.10) significance level.

BEST COPY AVAILABLE

American clients show less improvement in employment status. The finding is significant at the 0.10 level for the logistic model and at the 0.05 level for the ordered logistic model

On average, clients who present for treatment with more severe drug problems show less improvement in employment situation than do clients with less severe drug problems at intake. This effect, makes intuitive sense; however, it is not statistically significant in any of the three models.

Factors that have no statistically significant effect on change of employment situation include age, being Hispanic, and having more severe drug, family, legal, or alcohol problems at intake. Similarly, we find that severity of psychiatric problems at intake is not a significant predictor of an improvement in client employment outcomes, unlike the findings of Ouimette et al. (1999), Stahler et al. (1995) and Wright and Devine (1995) who find severity of psychiatric problems to be predictive of less improvement in employment outcomes.

Several studies find that employment at intake is a positive predictor of improved employment outcomes after treatment (Stahler et al., 1995; Wright and Devine, 1995; Lapham et al., 1995). Because we use change in employment situation as our dependent variable, our analysis explicitly controls for pre-treatment employment status. We include a measure of pre-treatment employment status solely to control for the statistical artifact that clients with better employment situations prior to treatment intake have less opportunity to improve upon their employment situation.

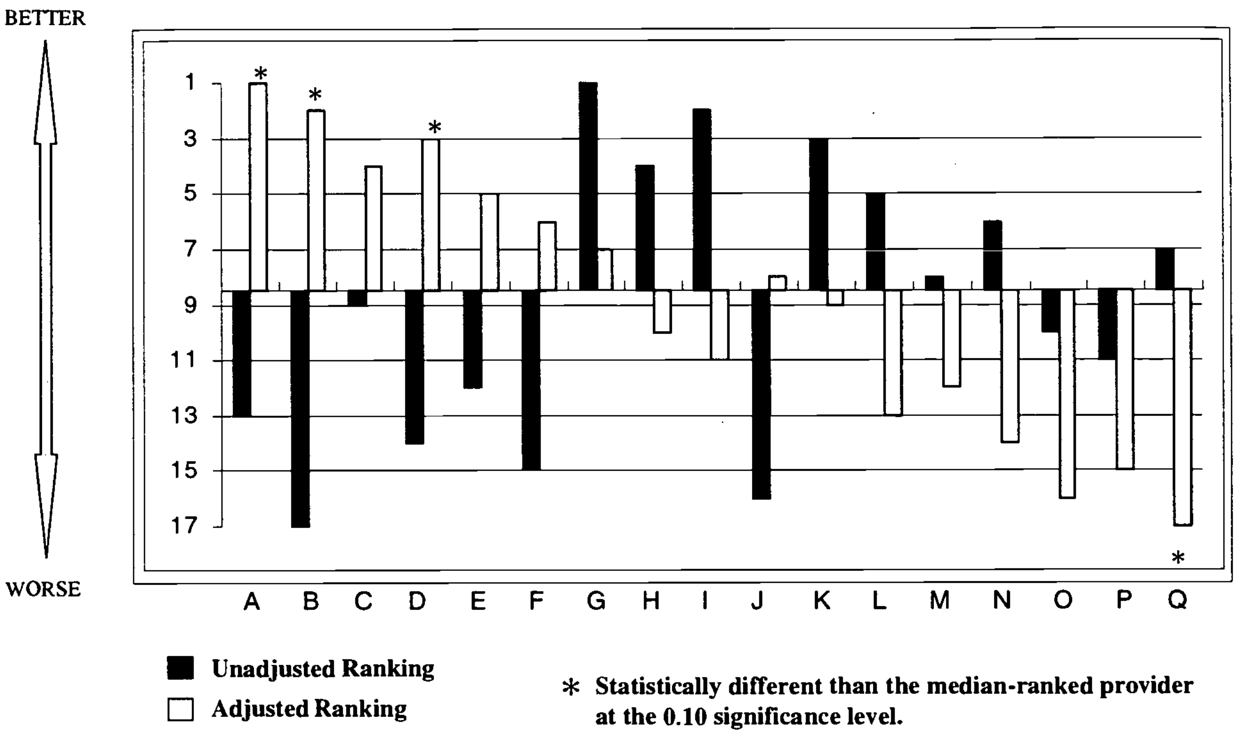
3. PROVIDER RANKINGS

We rank the 17 providers (including the composite provider, F) based on their estimated effectiveness relative to the effectiveness of the median ranked provider. For the ordinary least squares model, we used a t-test to identify providers whose estimated effectiveness is statistically different from the median-ranked provider's estimated effectiveness after controlling for case mix. We use a similar approach for the logistic and ordered logistic models, but use the Wald Chi-Squared statistic instead of the t-statistic.

3.1 Adjusted versus Unadjusted Rankings

Exhibit VI-3 presents the adjusted and unadjusted rankings for the 17 providers for each of the three models. As illustrated by other researchers, these results confirm the importance of using case-mix adjustment methods when comparing provider effectiveness. In the ordinary

EXHIBIT VI-4
ADJUSTED AND UNADJUSTED
OUTPATIENT RANKINGS: OLS MODEL



3.2. Consistency in Rankings Across Models

We find that the provider rankings are consistent across the *three case-mix adjusted models*. The rank of 7 providers (41 percent) does not vary by more than one place across the three models, and only 4 providers (24 percent) vary in rank by 4 or more places. We estimate a Spearman's Rank Correlation Coefficient (t) comparing the sets of rankings and calculate t=0.894 for the comparison of the rankings obtained by the ordinary least squares and logistic models, t=0.953 for the rankings from the ordinary least squares and ordered logistic models, and t=0.917 for the rankings from the logistic and ordered logistic models. These findings confirm the consistency in the rankings across the three models and should offer some comfort to providers and evaluators. However, it is important to recognize that, although the rankings remain fairly stable, the three models identified different providers as being statistically different from the median-ranked provider. If the purpose of the analysis is to identify whether or not a provider or set of providers is different (statistically) from a reference provider, then the way in which the outcome is measured and modeled appears to be an important consideration.

VII. SUMMARY AND IMPLICATIONS FOR RESEARCH, POLICY, AND PRACTICE

VII. Summary and Implications for Research, Policy, and Practice

1. SUMMARY

In this report, we demonstrated three applications of case-mix methods using regression analysis and used our results to assess the relative effectiveness of substance abuse treatment providers. We examined the ability of providers to improve client employment outcomes, an outcome domain relatively unexamined in the assessment of provider effectiveness. This outcome was measured as the change between the number of days clients were paid for work in the 30 days prior to the intake interview and the 30 days before the follow-up interview.

Consistent with previous research, our results confirm the need to use case-mix adjustment methods when assessing provider effectiveness. Although researchers may have long been aware of this finding, it is now crucial that Federal agencies, states, treatment providers and other entities currently, or soon to be, involved in the assessment of treatment providers also recognize the importance of case-mix adjustment. Analyses that account for difference in client characteristics reduce the risk of drawing inappropriate conclusions regarding the effectiveness of substance abuse treatment and, thus, limit the possibility that incorrect treatment and treatment funding decisions are made.

In addition, we found that our estimates of provider rankings varied little across the three regression models when controlling for case mix, suggesting that provider rankings are not especially sensitive to choice of method. While this may be of some comfort, its importance may be limited because the different models identified different sets of providers who differed statistically from the median-ranked provider. The ordinary least squares and ordered logistic model appear to be superior to the logistic model in this sense, since they were able to detect differences between providers that did not appear from the results of the logistic regression. This is, perhaps, not surprising considering that the ordinary least squares and ordered logistic models took into account more information on the change in days paid. However, it demonstrates the need for evaluators to craft outcome measures that reflect the most information, because that information may influence the estimates of provider effectiveness.

As a final thought, we believe that care needs to be exercised in how policy makers use case-mix adjustment. While case-mix analysis can be a valuable tool for assessing treatment provider effectiveness, it is important to recognize that such approaches have their limitations. Therefore, we see these approaches as a first step towards improving care. They can be used to identify providers who may require further examination. Through discussion with providers and additional study, one may be able to verify the case-mix findings and identify best practices for different client types. If applied appropriately, we believe these techniques can be useful to stakeholders and the provider community.

2. IMPLICATIONS FOR RESEARCH

We (and others) see a number of useful possible extensions to the analysis done for this report. First, it would be useful to test other employment-related outcome measures in order to validate the fundamental finding of this report that very similar results were obtained using different employment measures and methods. Several examples are to differentiate between those individuals working full time or part time, or to analyze the "quality" of work in terms of stability, pay, benefits, and opportunities. Our ability to work with such outcome measures was restricted because of data limitations. Another measure of employment that could be used to assess provider effectiveness is whether clients undergo unemployment spells during a specified period of time following treatment, and the length of those spells. Unemployment spells could be analyzed using survival analysis, which is a class of statistical methods used to analyze the occurrence and timing of events. Other names for survival analysis include "event history analysis," "duration analysis" and "transition analysis." This class of statistical methods has evolved largely from biomedical research, but has become increasingly common in the social sciences. By using these methods, more informative measures that take into account the type of employment and its duration should improve the usefulness of case-mix analysis of substance abuse treatment.

Equally important, while we were able to identify several "outlying" providers (applying statistical criteria), we were unable to determine why these providers differ from the median-ranked provider due to data limitations. Two main sources of variation resulting in differential effectiveness are differences in therapeutic approach and different structural features of the providers, neither of which are contained in the database we analyzed. Structural features of programs such as size, organization, staffing patterns, and other organizational characteristics may affect clients' outcomes. Furthermore, given the limitations of making policy decisions based on a single outcome domain (i.e., use of alcohol or drugs, employment outcomes, criminal activity), additional research is needed to develop efficient methods of assessing providers in terms of multiple outcome domains. The large body of literature on scale construction appears to hold promise for contributing to this need. Research that explores different weighting schemes for the different outcome domains will be of particular importance to policymakers who seek global measures of provider effectiveness that can enhance their contracting processes.

Finally, we believe that this report offers insights into how to design data systems and studies so that case mix analysis can be performed. Very few treatment effectiveness studies in the past have attempted or even considered making performance rankings of providers. However, this has become an important issue in recent years, and going forward it will be both possible and necessary to design studies in order to do case mix adjusted performance measurement.

3. IMPLICATIONS FOR POLICY

While this analysis has been fairly technical, we believe that there are important implications of this work. This analysis contributes to building a case for helping treatment systems and providers to be accountable for their performance. Both consumers and those paying for treatment (public agencies and private insurance) want to know that more effective and less effective providers can be identified, in order to learn from the former and improve the latter. Providers have long opposed such comparisons on the grounds that the patients served by different providers are quite different, and that rankings would therefore be inappropriate.

The case mix techniques applied in this analysis (and used successfully in other analyses) validate the concerns of providers at the same time, in that they demonstrate that it is possible to methodically "level the playing field" and generate "adjusted" (and therefor appropriate) performance rankings of substance abuse treatment providers. The fact that quite similar rankings/conclusions were generated using the various outcome measures and their appropriate analytic methods indicates that the methodology is "robust" and should yield similar results under modest variations.

However, we believe that policy makers should always confirm the conclusions from case mix adjusted performance rankings with direct information. Even when a "strong" case mix model is developed there is usually a lot that is unexplained, and managers, staff and clients can often provide invaluable insights that either validate or explain strong or weak rankings–information that reveals what works or doesn't work, for whom, and how to improve services.

4. IMPLICATIONS FOR PRACTICE

This analysis validates the concern of providers that different clients have different expected or predicted outcomes, and providers with more difficult clients need to be viewed differently than those with less severe clients. This is exactly what case mix adjustment is designed to address. Providers should be aware that performance measurement efforts are gaining momentum and they need to engage in the process by which these measurement systems are being developed and implemented in order to inform and shape the process and the system. We believe that case mix analysis is most meaningful when done together with site visits and case studies that yield meaning and give insight into the statistical analysis, and experienced providers are often among those best qualified to perform this service.

Providers could also use case mix analysis (given a meaningful data set) to monitor their own performance. This would allow them to identify when their performance was apparently improving or slipping in order to learn how to deliver services more effectively

REFERENCES

REFERENCES

Alterman, I.A., Brown, L.S., Zaballero, A., & Mckay, J.R. (1994). Interviewer Severity Ratings and Composite Scores of the ASI: A Further Look, *Drug and Alcohol Dependence*, 34, 201-209.

Ameen, A.Z., Dall, T.M., Harwood, H.J., & Doyon, V.A. (1999). *Comparisons of Substance Abuse Provider Effectiveness: An Analysis Using Case-Mix Adjustment* (manuscript under review).

Burnam, M.A. (1996). Measuring Outcomes of Care for Substance Use and Mental Disorders, *New Directions for Mental Health Services* 71, pp. 3-17.

Campbell, W.G. (1997). Evaluation of a Residential Program Using the Addiction Severity Index and Stages of Change, *Journal of Addictive Diseases*, 16 (2), 27-39.

Conrad, K.J., Hultman, C.I., Pope, A.R., Lyons, J.S., Baxter, W.C., Daghestani, A.N., McCarthy, M. Jr., & Manheim, L.M. (1998). Case Managed Residential Care for Homeless Addicted Veterans, *Medical Care*, 36:1, 40-53.

Dall, T.M., Ameen, A.Z., Koenig, L., & Harwood, H.J (1999). *Approaches for Making Risk-Adjusted Comparisons of the Performance of Substance Abuse Treatment Providers* (Draft manuscript). The Lewin Group, VA.

Erickson, J.R., Stevens, S., McKnight, P., & Figueredo, A.J. (1995). Willingness for Treatment as a Predictor of Retention and Outcomes, *Journal of Addictive Diseases*, 14(4), 135-150.

Greene, William H. (1997). *Econometric Analysis*. Upper Saddle River: Prentice-Hall.

Hannan, E., Kilburn, H., Bernard, H., O'Donnell, J., Lucacik, G., & Shields, E., (1991). Coronary Artery Bypass Surgery: The Relationship Between In-hospital Mortality Rate and Surgical Volume After Controlling for Clinical Risk Factors, *Medical Care*, 29(11), 1094-1107.

Hannan, E., Kilburn, H., Racz, M., Shields, M., & Chassin, M. A., (1994). Improving the Outcomes of Coronary Bypass Surgery in New York State, *Journal of the American Medical Association*, 271(10), 761-766.

Hannan, E., Siu, A., L., Kumar, D., Kilburn, H., & Chassin, M. A. (1995). The Decline in Coronary Artery Bypass Graft Surgery Mortality in New York State, *Journal of the American Medical Association*, 273(3), 209-213.

Harwood, H.J., Bazron, B., & Fountain, D. (1997). *Performance Measurement for Substance Abuse Treatment Providers and CSAT Knowledge, Development, and Application.* A report to the Center for Substance Abuse Treatment, Rockville, MD.

Lapham, S.C., Hall, M., & Skipper, B.J. (1995). Homelessness and Substance Use Among Alcohol Abusers Following Participation in Project H&ART, *Journal of Addictive Diseases: the official Journal of the ASAM*, 14:4, 41-55.

Leda, C., & Rosenheck, R. (1992). Mental Health Status and Community Adjustment After Treatment in a Residential Treatment Program for Homeless Veterans, *American Journal of Psychiatry,* 149:9 1219-1224.

Luft, H.S., & Romano, P.S. (1993). Chance, Continuity, and Change in Hospital Mortality Rates, *Journal of the American Medical Association*, 270(3), 331-337.

Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics,* Cambridge University Press.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The Fifth Edition of the Addiction Severity Index, *Journal of Substance Abuse Treatment,* 9, 199-213.

McLellan, A.T., Grant, G., Brill, P., Durell, J., Metzger, D., & O'Brien, C. (1993). Private Substance Abuse Treatments: Are Some Programs More Effective Than Others? *Journal of Substance Abuse Treatment*, 10, 243-254.

Ouimette, P.C., Gima, K., Rudolf, H.M., & Finney, J.W. (1999). A Comparative Evaluation of Substance Abuse Treatment IV. The Effect of Comorbid Psychiatric Diagnoses on Amount of Treatment, Continuing Care, and 1-Year Outcomes, *Alcoholism: Clinical and Experimental Research,* 23:3, 552-557.

Prochaska, J.O. & Diclemente, C.C. (1986). Toward a Comprehensive Model of Change, In W.R. Miller & N. Heather (Eds.*), Treating Addictive Behaviors: Process of Change (pp. 3-27)* (New York: Plenum Press).

Prochaska, J.O. & Diclemente, C.C. (1992). Stages of Change in the Modification of Problem Behaviors. In M. Hersen, R.M. Eisler, and P.M. Miller (Eds.), *Progress in Behavior Modification* (pp. 184-214), (Sycamore, IL: Sycamore Press).

Phibbs, C.S., Swindle, R.W., & Recine, B. (1997). Does Case Mix Matter for Substance Abuse Treatment? A Comparison of Observed and Case Mix-Adjusted Readmission Rates for Inpatient Substance Abuse Treatment in the Department of Veterans Affairs, *Health Services Research*, 31(6), 755-771.

Phillips, C.D., Hubbard, R., Dunteman, G., Fountain, D., Czechowicz, D., & Cooper, J.R. (1995). Measuring Program Performance in Methadone Treatment Using in-Treatment Outcomes: An Illustration, *The Journal of Mental Health Administration*, 22(3), 214-225.

Smith, D.W., McFall, S.L., & Pine, M.B. (1993). State Rate Regulation and Inpatient Mortality Rates, *Inquiry*, 30, 23.

Stahler, G.J., Shipley, T.F., Bartlet, D., Ducette, J.P., & Shandler, I.W. (1995). Evaluating Alternative Treatments for Homeless Substance-Abusing Men: Outcomes and Predictors of Success, *Journal of Addictive Diseases: The Official Journal of the ASAM,* 14:4, 151-167.

Velicer, F.W., Hughes, S.L., Fava, J.L., Prochaska, J.O., & Diclemente, C.C. (1995). An Empirical Typology of Subjects within Stage of Change, *Addictive Behaviors*, 20(3), 299-320.

Wright, J. & Devine, J. (1995). Factors That Interact with Treatment to Predict Outcomes in Substance Abuse Programs for the Homeless, *Journal of Addictive Diseases: The Official Journal of the ASAM,* 14:4, 169-181