

# INTEGRATED EVALUATION METHODS



## A GUIDE TO SUBSTANCE ABUSE TREATMENT EVALUATION DATA ANALYSIS

July 1999

**NEDTAC**

**CSAT**  
Center for Substance  
Abuse Treatment  
SAMHSA

**CALIBER**  
ASSOCIATES  
Battelle

# INTEGRATED EVALUATION METHODS



## A GUIDE TO SUBSTANCE ABUSE TREATMENT EVALUATION DATA ANALYSIS

**Prepared by:**

**Douglas Fountain  
Caliber Associates**

**Rob Orwin, Ph.D.  
and  
Michael Maranda, Ph.D.  
Battelle**

**July 1999**

This document was produced by the Center for Substance Abuse Treatment, Department of Health and Human Services, Caliber/NEDTAC Contract No. 270-94-0001 and is being made available through Caliber/NEDS Contract No. 270-97-7016.

**CSAT**  
Center for Substance  
Abuse Treatment  
SAMHSA

# TABLE OF CONTENTS

	<u>Page</u>
<b>FOREWORD</b> .....	<b>i</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>iii</b>
<b>I. INTRODUCTION</b> .....	<b>1</b>
1. PURPOSE OF THIS DOCUMENT .....	1
2. INTEGRATIVE EVALUATION METHOD TOOLS .....	2
<b>II. OVERVIEW OF EVALUATION QUESTIONS AND ANALYTIC STRATEGIES</b> .....	<b>4</b>
1. CATEGORICAL FRAMEWORK FOR QUESTIONS AND ANALYSES .....	4
2. STRATEGIES TO ACHIEVE STRONG ANALYTIC RESULTS .....	8
2.1 Integrate Process and Outcome Information .....	9
2.2 Analyze Quantitative Data .....	10
2.3 Analyze Qualitative Data and Integrate with Quantitative Findings .....	13
2.4 Complete Analyses Across Multiple Sites .....	14
3. DEVELOPING CONCLUSIONS AND RECOMMENDATIONS .....	15
<b>III. SPECIFIC QUESTIONS AND ANALYTIC APPROACHES</b> .....	<b>18</b>
1. DESCRIBING THE PLANNED SERVICE .....	18
2. ASSESSING SERVICE DELIVERY .....	27
3. DESCRIBING IMPACTS .....	44
4. APPLYING GOOD JUDGMENT IN THE USE OF DATA .....	65
4.1 Assess Reliability and Validity .....	65
4.2 Review Data Collection Procedures and Assess Data Quality .....	68
4.3 Assess the Evaluation Design .....	69
<b>REFERENCES</b> .....	<b>77</b>

## TABLE OF CONTENTS (CONTINUED)

	<u>Page</u>
<b>APPENDIX A: INTEGRATED EVALUATION METHODS PACKAGE: A GUIDE FOR SUBSTANCE ABUSE TREATMENT KNOWLEDGE- GENERATING ACTIVITIES—EXECUTIVE SUMMARY .....</b>	<b>79</b>
<b>APPENDIX B: MEASUREMENT CONCEPTS .....</b>	<b>85</b>

## FOREWORD

Over the last 10 years, the Center for Substance Abuse Treatment (CSAT) has accumulated a great deal of experience in substance abuse treatment evaluation implemented through coordinating centers, cross-site efforts, and national studies. The importance and value of integrating ongoing evaluation activity into a system for treating substance abuse problems is widely recognized. Also widely recognized, however, is that current evaluation-generated knowledge and practice are often under-utilized, due in part to the lack of an integrated approach to capturing information with which to measure and improve treatment effectiveness, efficiency, and performance. CSAT recognizes that such an integrated evaluation approach will more effectively support current and future knowledge-generating activities.

Based on a decade of evaluation experience, CSAT has developed the Integrated Evaluation Methods (IEM) Package, a series of conceptual and methodological applications, including concept papers, technical assistance materials, and analytic tools, to enhance CSAT-funded evaluation activities. Products in the IEM Package are organized within an evaluation framework constructed on the basis of accumulated experiences among internationally known treatment service evaluation professionals. Thus, the framework is based upon evaluation strategies, structures, and approaches appropriate for substance abuse treatment evaluators and providers. The framework follows a standard set of evaluation activities: planning, selecting a design, developing data requirements and collection instruments, collecting and analyzing the data, and reporting the evaluation findings. (A summary description of the IEM Package is contained in Appendix A to this document.)

This analytic tool, together with its companion documents *A Guide to Process Evaluation for Substance Abuse Treatment Services*, *Using Logic Models in Substance Abuse Treatment Evaluations*, and *A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations* is aimed at supporting the analysis phase of the evaluation. Specifically, this paper builds upon a framework of common data collection methods and core data lists published in two additional products—the *Minimum Evaluation Data Set (MEDS)* and the *Substance Abuse Treatment Cost Analysis and Allocation Template (SATCAAT)*—to provide a tool for combining these data with process data into an integrated analysis plan.

Sharon Bishop  
Project Director  
NEDTAC

## ACKNOWLEDGMENTS

This paper was developed for CSAT by the National Evaluation Data and Technical Assistance Center (NEDTAC) under the guidance and direction of Ron Smith, Ph.D., Program Evaluation Branch, Office of Evaluation, Scientific Analysis, and Synthesis (OESAS). Caliber Associates was the prime contractor for NEDTAC in partnership with Computech, Inc.; the Lewin Group; Capital Consulting Corporation; the Center for Substance Abuse Research (CESAR), University of Maryland; the Alcohol Research Group (ARG), Public Health Institute; and the Drug Abuse Research Center (DARC), University of California, Los Angeles; and the Urban Institute.

NEDTAC is extremely grateful to Doug Fountain, who is the principal author of the paper and to Rob Orwin, Ph.D., and Michael Maranda, Ph.D., who made major contributions to the discussion of and guidance for outcome analysis. In addition, several people within Caliber contributed to this effort. Patricia Devine, Marsha Morahan, Robin Walthour, Erica Sorohan, and Donna Caudill provided essential contributions. Contributions ranged from report editing to report production, and all of these efforts were equally invaluable and greatly appreciated.

# I. INTRODUCTION

The Center for Substance Abuse Treatment (CSAT), using the resources of the National Evaluation Data and Technical Assistance Center (NEDTAC) contract, developed a series of methodological concept papers and technical assistance materials to enhance and integrate CSAT-funded evaluation activities. These evaluation activities included grant-specific “site-specific” evaluations as well as program-level evaluations implemented through coordinating centers, cross-site efforts, and national studies.

Based on the site-specific, cross-site, and national evaluation experiences, CSAT recognizes the need for a uniform, integrated set of evaluation methods and technical assistance materials. This document is one component of the technical assistance materials designed for use in all knowledge-generating evaluation activities sponsored by CSAT that can be used by evaluators throughout the substance abuse treatment research and evaluation fields. National and site-specific evaluators can use this information to gain an understanding of the overall CSAT evaluation strategies. This understanding, together with the integrated evaluation methods and technical assistance materials, will assist in planning and conducting national evaluation studies, multi-site and cross-site efforts, and site-specific evaluations.

This summary of analytic strategies and approaches is necessary because the analysis, like most other parts of an evaluation effort, needs to be carefully planned. For too many evaluations, tight timelines and limited budgets cause the analysis to be curtailed after what is often years of effort spent designing, collecting, and reviewing information. In addition, treatment service managers become concerned about how evaluations will portray the treatment services at which they have toiled for years. For these reasons, it is in the evaluator’s best interest to plan as thoughtfully as possible how information will be gathered and used, while avoiding the problems of limited resources and timelines. This analysis plan walks evaluators through the steps necessary to describe the objectives and results of their evaluation effort for clients, organizations, and other audiences.

## 1. PURPOSE OF THIS DOCUMENT

This document has assembled a series of questions, references to analytic tools, and a discussion of tips or techniques for completing an analysis of evaluation questions. Generally speaking, this material will serve as a guide for those who are designing, managing, or overseeing substance abuse treatment evaluation activities, including those sponsored by CSAT. The questions will guide evaluators through the decision-making process for planning and

executing an evaluation. CSAT staff, grant monitors, grant evaluators, and evaluation contractors can use these evaluation questions as a guide for standardizing the evaluation process and integrating results across evaluations. CSAT program monitors and treatment service directors can use these questions as a framework for designing knowledge-generating programs and studies, for defining expectations of an evaluation, and for monitoring adherence to those expectations. Evaluators can check their own plans against this list of questions to ensure that at the end of the evaluation process, they have collected sufficient information to address these issues.

This document is designed to be used both as a planning guide and a checklist during the analysis phase. As a planning guide, the evaluator should take steps to ensure that most of the evaluation questions presented herein can be addressed. As a checklist, the evaluator can use this document to systematically move through both the conceptual and the practical aspects of assembling and reviewing data and drawing conclusions to inform future policies or actions.

## **2. INTEGRATIVE EVALUATION METHOD TOOLS**

CSAT's evaluation experiences have reinforced the fact that substance abuse treatment evaluation involves a standard set of tasks that generally occur in the following order:

- **Planning the evaluation**, which includes setting the evaluation goals and objectives that determine the overall parameters of the evaluation
- **Selecting the evaluation design**, which sets forth the overall strategy for establishing the evaluation questions, measurement approach, and generalizability of findings
- **Developing the data requirements**, which flow from the evaluation questions and measures and include: service delivery unit (SDU), clinician, cost, and client data
- **Developing data collection instruments**, which are based on the data requirements and are developed or selected from a standard inventory of instrumentation
- **Collecting the data**, which includes the development of data management processes and tools including quality control procedures, and collecting the data
- **Analyzing the data**, which usually involves multiple levels of comparison and is governed by an analysis plan and intended products and target audience(s)
- **Reporting the evaluation findings**, which includes evaluation knowledge dissemination and application within the field.

CSAT has directed the development of evaluation concepts, tools, and technical assistance materials to support these evaluation tasks that are summarized in Exhibit I in Appendix A. The document, *A Guide to Substance Abuse Treatment Evaluation Data Analysis*, has been designed to support the sixth stage in the evaluation process, Reporting the Evaluation Findings. A full discussion of the CSAT evaluation framework and the other evaluation concepts, tools, and technical assistance materials is presented in the concept paper *Integrated Evaluation Methods (IEM): A Guide for Substance Abuse Treatment Knowledge-Generating Activities*.

This document provides a tool for combining process evaluation variables, outcome evaluation variables (e.g., *Minimum Evaluation Data Set* data), and cost analysis variables (e.g., *Substance Abuse Treatment Cost Analysis and Allocation Template*, or SATCAAT variables) into an integrated analysis plan.

The next chapter presents a framework of information categories that need to be addressed by most treatment evaluations, as well as general and specific suggestions on the process of combining and analyzing data sources to generate evaluation conclusions. Chapter III provides a breakdown of evaluation questions and analytic strategies for the following topics: describing planned services, assessing service delivery, assessing service impacts, and using good judgment in evaluation findings.

The IEM document and the associated evaluation concept papers and technical assistance materials are fully referenced in Appendix A. A discussion of measurement concepts is presented in Appendix B.

## II. EVALUATION QUESTIONS AND ANALYTIC STRATEGIES

In its simplest form, the analysis of evaluation data amounts to articulating a series of questions and then bringing to bear all possible valid and reliable information to help shed light on those questions. This chapter presents a categorical framework for evaluation information and summarizes literature and practical experience on the analysis and utilization of evaluation findings. The emphasis is on identifying “crosscutting” issues and themes that apply in many situations, including:

- Integrating process and outcome data
- Using both high-quality quantitative and qualitative data
- Conducting multi-site studies
- Drawing conclusions and recommendations.

This paper presumes that an analyst has an underlying set of client, counselor, SDU, and cost variables in place but balances that against the need to not be bound to existing variable sets when attempting to answer complex questions. On one hand, this paper presents clear guidance for analysts who have “canned data” from the *Minimum Evaluation Data Set (MEDS)* and the *Substance Abuse Treatment Cost Analysis and Allocation Template (SATCAAT)* products, and process information similar to that described in the *Guide to Process Evaluation for Substance Abuse Treatment Services*. However, for any given evaluation, the guidance offered here should be tailored and specified. Chapter III offers additional information to help the reader move from these general issues to more specific evaluation questions and analytic strategies. To assist the analyst in this task, the next section presents a framework for performing analysis.

### 1. CATEGORICAL FRAMEWORK FOR QUESTIONS AND ANALYSES

Underlying the analysis of substance abuse treatment evaluation data is the need to clearly articulate a set of specific and well-constructed evaluation questions and then to ensure that these questions are answered using the best possible evaluation design with valid and reliable measurements. Evaluators of substance abuse treatment services should begin developing specific evaluation questions by considering broad questions about the initiative being studied. This document relies on the following framework for exploring substance abuse treatment service designs, services, and outcomes:

- What are the goals of the service, and how is it designed to be different from other services?
- What services are actually delivered, to whom, when, and at what cost?
- What were the impacts of the services, and the relative cost of achieving those impacts?

Evaluators recognize these overarching questions as corresponding to the major stages of an evaluation: goals, process, outcomes, and cost-effectiveness. To facilitate the integration of results, each evaluation of a substance abuse treatment initiative should address a standard set of information categories. These categories are: (1) describing the planned services, (2) assessing service delivery, and (3) describing the impacts. A fourth category emerges as a context for the prior three: (4) considering factors that bolster or weaken the strength of the evaluation findings. Exhibit II-1 presents a set of topics within these categories that the evaluator should seek to address. Identifying these categories is a first step toward preparing specific evaluation questions.

In developing evaluation questions, the investigator needs to consider the purpose or purposes of the evaluation, i.e., what the evaluation is trying to accomplish. Chelimsky (1978)

<b>EXHIBIT II-1 EVALUATION INFORMATION CATEGORIES</b>		
<b>1. DESCRIBE PLANNED SERVICE</b>	<b>2. ASSESS SERVICE DELIVERY</b>	<b>3. DESCRIBE THE IMPACTS</b>
1.1 Development of service	2.1 Implementation process	3.1 Retention and participation
1.2 Gaps and goals	2.2 Clients served	3.2 Alcohol and drug abuse
1.3 Logic model	2.3 Treatment process	3.3 Social and family outcomes
1.4 Planned changes in process	2.4 Staff characteristics	3.4 Employment and productivity
	2.5 SDU characteristics	3.5 Health
	2.6 External environment	3.6 Mental health
	2.7 Cost of services	3.7 Criminal justice
		3.8 Economic benefits and costs
<b>4. ASSESS FACTORS THAT BOLSTER OR WEAKEN THE EVALUATION FINDINGS</b>		
4.1 Evaluation Design		
4.2 Measurement strategies and data quality		
4.3 Analytic approaches and methods		
4.4 Comparison of results with literature		

identified three principal purposes for evaluation: more meaningful accountability, improved program delivery, and knowledge generation. While both effort and effectiveness questions support all three purposes to a degree, knowledge development is most often associated with effectiveness questions. Its principal intent is the development of valid, generalizable knowledge about interventions, the social problems on which interventions are targeted, the social systems within which interventions are implemented, and the effects of those interventions on targeted outcomes.

In this document, we rely on two approaches to evaluation of substance abuse treatment services: (a) comparisons of two or more existing services in order to shed light on which service may be more effective; and (b) comparisons of the results of introducing a change in a treatment service. These approaches are more similar than dissimilar: Both entail comparisons of treatment goals, processes, and outcomes across two or more treatment conditions. The questions, however, need to be tailored to the specific context.

- **Comparisons of Two Existing Conditions.** Any evaluation that compares existing services and outcomes across multiple service settings needs to demonstrate the characteristics of groups receiving services and of the services those groups receive, focusing on similarities and differences across groups and the veracity of the factors believed to be different across treatment conditions. For example, a comparison of an 8-week treatment regimen with a 16-week treatment regimen needs to determine whether clients in fact stayed an average of 8 weeks in the first and 16 weeks in the latter.
- **Comparisons of Changed Services with a Standard Condition.** Questions should ultimately drive the evaluator to consider how processes and outcomes are truly changed relative to a standard or comparison condition. Many evaluators collect data from a comparison group or pre-implementation baseline or both. Along with this, the evaluator of a change-based initiative must consider how the intervention “matured” or “adapted” over time, and how that affected the costs and results of the intervention. Interventions mature as they become more routine and staff gain experience. Staff may adapt the process to fit emerging circumstances in ways that affect the ability to achieve predicted results. As interventions evolve, evaluators may find that the analysis of processes and results will look quite different early on compared to later after implementation.

When articulating the “main effects” or central findings from either of the above types of evaluation, the evaluator is also responsible for ensuring that evaluation designs and measurement strategies do not create or mask evaluation results. For example, if the evaluation requires a comprehensive, standardized client needs assessment in order to measure effects of an

intervention, that comprehensive assessment may itself lead to improved outcomes. This effect could in turn mask—or bolster—the apparent effects of the treatment initiative. The evaluator should consider this perspective and ensure that the evaluation will control for unintended effects.

<b>EXHIBIT II-2</b> <b>EXAMPLES OF APPLYING PRIMARY EVALUATION QUESTIONS</b> <b>TO TWO TYPES OF STUDIES</b>		
<b>INFORMATION CATEGORIES</b>	<b>COMPARISONS OF TWO EXISTING CONDITIONS</b>	<b>COMPARISONS OF CHANGED SERVICES WITH A STANDARD CONDITION</b>
<b>1. Describe the planned service</b>	What are respective goals across treatment conditions?	What is the goal for the change?
<b>2. Assess service delivery</b>	How many services, of what type, and for what duration did clients get across conditions?	How did services actually change, compared to baseline information, comparison group, or both?
<b>3. Describe the results</b>	What are the relative outcomes observed between conditions?	Were the outcomes of “matured” service changes better than baseline or comparison group?
<b>4. Analyze how evaluation process relates to findings</b>	Are data collection methods consistent across different service delivery units?	Are data collection methods part of the intervention that could change apparent differences between groups?

Exhibit II-2 demonstrates how to begin to tailor questions to address fundamentally different designs. For questions to be useful in an evaluation study, they must be precisely defined for the study. This usually means framing each question in measurable terms. An abstract notion needs to be turned into a measurable reality in order to be answered. The manner in which outcomes are *measured*, therefore, is a critically important part of any outcome evaluation. A poorly conceptualized measure can completely undermine the value of the evaluation by producing misleading estimates of substance abuse treatment effectiveness.

The evaluator must make evaluation questions as specific as possible to help ensure that the most appropriate data are identified and analyzed. For example, for a substance abuse treatment service that includes a job readiness component, the evaluator may ask how many clients were employed 6 months after completing the treatment services, and how their employment status compared with their previous employment history or with employment outcomes among a comparison group. In order to answer these questions, the evaluator must define terms. How is employment defined? Does legal, full-time employment count, or does

“off-the-books” and part-time employment count as well? How is previous employment measured? Over what time frame is employment measured?

Not all knowledge-generating evaluations have immediate practical impacts, but they may lay the foundations for far-reaching substance abuse treatment innovations in the longer term. Also, knowledge-generating evaluations must remain flexible to cope with an environment that includes changing policies (e.g., the switch to managed behavioral health care) and changing social problems (e.g., the resurgence of methamphetamine abuse in the United States). The Substance Abuse and Mental Health Services Administration’s (SAMHSA’s) Knowledge Development and Application (KD&A) programs provide excellent examples of knowledge development evaluations that collectively serve both purposes.

## **2. STRATEGIES TO ACHIEVE STRONG ANALYTIC RESULTS**

Upon completion of the data collection task, the evaluation team will have acquired a wealth of interview, objective, and observational data. These data need to be sorted, compiled, aggregated, analyzed, and compared to answer key evaluation questions. Key steps in the analytic process include:

- Organize and aggregate available data by type (interview, objective, and observation) and by evaluation area (design, implementation, treatment service component descriptions, and factors affecting outcomes).
- Analyze each data set and develop findings for each “condition” being studied.
- Compare and contrast findings across treatment conditions.
- Identify intervening variables or factors that may affect other findings.
- Develop overall findings for each treatment condition based on all types of data relevant to that condition.
- Consolidate and compare findings from different evaluation areas to develop an overview of the services and their impacts.
- Compare these overall findings with prior results from the literature.

The task for an evaluation analysis is to tell a story about a service or an initiative from beginning to end. As a story, it should have a theme and purpose and an audience in mind; the

story should weave together many pieces of quantitative and qualitative information for descriptive and inferential purposes. The following sections discuss the importance of integrating process and outcome information, analyzing quantitative and qualitative data, completing analyses across multiple sites, and drawing conclusions and recommendations.

## 2.1 Integrate Process and Outcome Information

The evaluation should pry open what is commonly referred to as the “black box” of treatment. The “black box” refers to the combination of structures (staff, facilities) and processes (therapeutic strategies, services) that constitute the treatment experience faced by one client or a group of clients. Anecdotally, treatment is like a pharmaceutical in that some aspects of treatment are “inert” and have no specific measurable effect, while other aspects are “active” ingredients. Active ingredients are those that have the most direct effect on recovery and relapse prevention. The evaluator must identify the active ingredients in treatment, and how those active ingredients resulted in demonstrated outcomes. This requires that the evaluator systematically and thoroughly think about how aspects of the treatment process influence outcomes.

It is easy for audiences of evaluation reports to focus only on outcomes if the evaluator does not probe the reasons for outcome findings. Data that demonstrate differences or lack of differences between groups being studied can become persuasive; however, the evaluator needs to turn to evidence of differences in populations, services, treatment service implementation, and contextual factors that can account for outcome differences. In fact, when outcome evaluations reveal little impact, it often “turns out that implementation is faulty or incomplete” (Rossi and Freeman, 1993).

To illustrate how process information goes hand in hand with outcome information, consider the analysis of a change-based evaluation. Comparisons between treatment conditions are spurious if the treatment innovation was not implemented. While this overstates the case, there are many “shades of gray” in determining whether a service was truly implemented. If only a fraction of the clients assigned to the service actually receive it, or if those that participate only attend a fifth of the total sessions, then the service perhaps never had the opportunity to generate improved outcomes. At the least, the improvements were likely mitigated. On the other hand, if the treatment service was not implemented well, but differences between groups *were* detected, the evaluator must pursue alternative explanations. The evaluator would be well served in such a situation to fully consider *why* the service was not fully implemented or attended.

## **2.2 Analyze Quantitative Data**

Data analysis requires several labor-intensive steps that include preparing appropriate data files for analysis, summarizing data, and performing statistical analysis as described below.

### **Clean the Data and Prepare Analysis Files**

Even with good data collection procedures, missing values and “bad data” are likely to occur. It is a problem the evaluator needs to handle appropriately. Fortunately, the analyses of data with missing values has progressed considerably since “plugging in the mean” was considered a reasonable response. The current state-of-the-art is a process called *multiple imputation*.

Leave missing and bad data codes out of analysis. Most statistical packages have a convention for excluding missing data; however, in some data sets, a missing value may take on a legitimate value. Some researchers key missing or bad data with out-of-range values such as “-9” or “99” for a categorical response. Be sure to change such values to missing before performing analysis. Many junior data analysts who do not know this may inadvertently leave these missing codes in the analyses, distorting the results.

Check also for data that are out of the possible range, for bad data, and for impossibly inconsistent data, and recode to missing. For example, range checks may show that the only client in an adult treatment service who is less than 19 years old is 3 years old. This is likely a miscoded response. A similar and classic example of inconsistent data is a “male” who indicates he is “pregnant.” One or the other coded response is wrong.

Be willing to use your judgment, and be sure to document your procedures. Data cleaning sometimes requires judgment on the part of the evaluator. For example, a distribution of annual earned legitimate income may show one or more respondents with annual incomes of over \$300,000. In publicly funded treatment providers, such income levels are highly uncommon and may suggest that an error was made in recording or keying the data; however, it is not impossible. Analyses with such a value could skew findings, and the evaluator will need to decide if the data are good or bad. The procedures for checking for bad data are discussed in “summarizing data” below. In all cases, data that are clearly bad should be removed from analysis and either left missing or replaced using the multiple imputation procedure mentioned above, and the decision rules employed for such recodes should be communicated.

## Summarize the Data

For all quantitative variables being studied (e.g., see specific analyses discussed in Chapter III), begin by carefully examining the properties of each separate variable being analyzed. Most commonly, measurements are summarized in terms of their central tendency and their dispersion or range. The type of measurement determines how central tendency and dispersion can be described. Summary statistics, such as SAS's PROC UNIVARIATE—which returns mean, median, mode, ranges, quintiles, and highest and lowest values for variables—should be used for non-categorical data. For categorical variables with a limited number of categories, a simple frequency distribution accomplishes the same objective. With a frequency distribution, analysts can verify that out of range, bad, and missing data have been left out of statistical calculations.

- **Calculate and analyze frequency distributions for nominal measures (e.g., categorical responses).** Nominal measurements have no central tendency because they are not ordered in any way. For example, gender and race/ethnicity variables have no logical ordering. Means and standard deviations are not appropriate for this type of data. However, reporting the frequency of occurrence of a specific category is necessary.
- **Calculate and analyze median, range, and quartile distributions or other ranking routines (e.g., quintiles) for categorical measurements that have a logical ranking.** Measurements that can be ordered from the lowest to highest rank can be organized to identify values that are greater than 25, 50, or 75 percent of the responses. Report the median value, below which 50 percent of observations fall, and the range from lowest to highest.
- **Calculate and analyze the central tendency with mean or arithmetic average along with standard deviations for interval and ratio data.** The mean uses each measurement's value in its calculation, not just its frequency or ordered position. Calculate the standard deviation to characterize dispersion of values around the mean. Determine whether the standard deviation is too large relative to the mean, rendering the mean less interpretable. Also, consider whether the mean is skewed by particularly large or small values in data sets with small sample sizes, making it less reflective of a "typical" response.

When the evaluator is comfortable that the data are free of bad and missing values and when the evaluator understands what single variables indicate about the population, it is time to select appropriate bivariate or multivariate statistical analyses.

## Perform Statistical Analyses

Perform statistical analyses to test hypotheses about the simple relationship between two variables (bivariate analysis) or the unique impact of a series of variables upon a single dependent variable (multivariate analysis). The exact statistical analyses performed will be contingent upon the type of data (e.g., categorical versus interval data), distributions of values for each variable, and the sample size. The evaluator needs to take into account the total number of variables he or she plans to analyze, recognizing that trying to subdivide the analysis data by too many groups (e.g., by including too many variables in a statistical model) potentially washes out main effects.

- **Consider the degree of association among pairs of measurements.** As with central tendency, the appropriate ways to describe association also depends upon the level of measurement. For a pair of interval or ratio-level measurements, calculate the Pearson Product-Moment Correlation Coefficient, or Pearson. For ordinal-level data, calculate the Spearman Rank Correlation Coefficient, noting, however, that this calculation is based on how the ranks of the measurements vary together rather than the values themselves. For non-ordered categorical data, calculate the phi coefficient, which is derived from a chi-square analysis generally of cross-tabulations of the variables. Remember that these tests assess association between variables, not necessarily causation.
- **Tease out the contribution of other variables from the treatment's effect on results using multivariate statistical techniques.** Multivariate techniques include multiple regression for outcome data that are continuous, or multiple logistic regression for dependent variables that are dichotomous. For example, a regression could be used to determine whether length of stay in treatment resulted in better change scores from pre- to post-treatment on drug use, controlling for other factors such as client characteristics. If more than one site is involved in the study, then a hierarchical model can be used to account for differences among the sites or other characteristics across the sites.

Be clear about underlying assumptions and at least discuss, if not test, how differences in the assumptions change ultimate findings. A common assumption is that non-respondents would have had data with similar properties as respondents, e.g. that the central tendency or relative frequency distributions would be unaltered. Sensitivity analyses should be performed, chiefly to demonstrate that assumptions made in the statistical analysis, if not correct, basically do not matter in the final analysis. For example, recoding non-respondents to demonstrate worst-case possible outcomes would allow a researcher to determine if treatment was still effective even

with the “worst case” for non-respondents. The point here is to question the assumptions and test, with sensitivity analysis, how they may relate to ultimate results.

### **2.3 Analyze Qualitative Data and Integrate with Quantitative Findings**

When exploring qualitative information sources, evaluators should seek to “triangulate” on preliminary conclusions before they are accepted as “fact.” Qualitative data are subject to the same threats as are quantitative data, such as construct validity, interrater reliability, and test-retest reliability. Often times, the perspective reflected by one or a few staff members, or one interoffice memo, or one stakeholder experience is just that: a perspective from only one or few sources. If multiple sources echo common themes, however, then the evaluator can have more confidence in the accuracy of that information.

Integrate findings from qualitative and quantitative sources. Do the numbers back up the experience? Does the experience back up the numbers? Qualitative information provides the context for judging or considering the quantitative findings; as such, quantitative results may support qualitative insights. Here is an example.

Staff in the XYZ Agency use a new treatment guideline while the ABC Agency uses old “tried and true” approaches. In comparing results between the ABC Agency and the XYZ Agency, an evaluator discovers that clients in both groups received the same number of services but XYZ’s outcomes are better than ABC’s. While tempted to say that the new guideline improved treatment, the evaluator may explore qualitative insights too. In ABC, staff are less professionally experienced than at XYZ, and there have been a lot of problems in the agency. Clients engage better at XYZ than at ABC, which may reflect the differences in staff and service environment. The evaluator could then draw the conclusion that observed differences in outcomes could partially be mitigated by exogenous influences in ABC.

Before integrating qualitative and quantitative results, however, explore the quality of each type of data.

### **2.4 Complete Analyses Across Multiple Sites**

With completion of analysis of the individual evaluations for individual services, evaluators involved in multi-site studies should turn to aggregate and comparative analyses that incorporate specific findings obtained from the individual site analyses. The aggregate and

comparative analysis of treatment services across sites is the definitive analytic step for any given multi-site evaluation. In this step, the findings from each evaluation area are compared across sites for the following purposes:

- To develop an overview of the evaluation area and the range of experiences or results that were obtained
- To identify similarities or differences in implementation
- To attempt to account for the differences that occurred and the reasons why they occurred
- To develop an aggregate assessment of the services being studied, including their overall viability and implications from comparative analysis for sustainability and replicability.

Multi-site analyses must incorporate process, outcome, and cost data. Process data allow the analysts to understand similarities and differences in the treatment services, or factors that drive differences in implementation of new initiatives. These may directly explain differences in outcomes that are obtained.

Moreover, analyze outcome data so as to control for the influence of specific sites. For example, in regression analyses, the evaluator should test models that include dummy variables for all but one site (a random “control site” selected to avoid the dummy variable trap). When specific sites become significant predictors of a dependent variable, the evaluator should explore further why outcomes are particularly high or low at that site. The evaluator should compare the impact of including site identifiers on other independent variables by running the model with and without the site dummies. The significance and magnitude of predictive variables may be mitigated or enhanced by inclusion of site dummies.

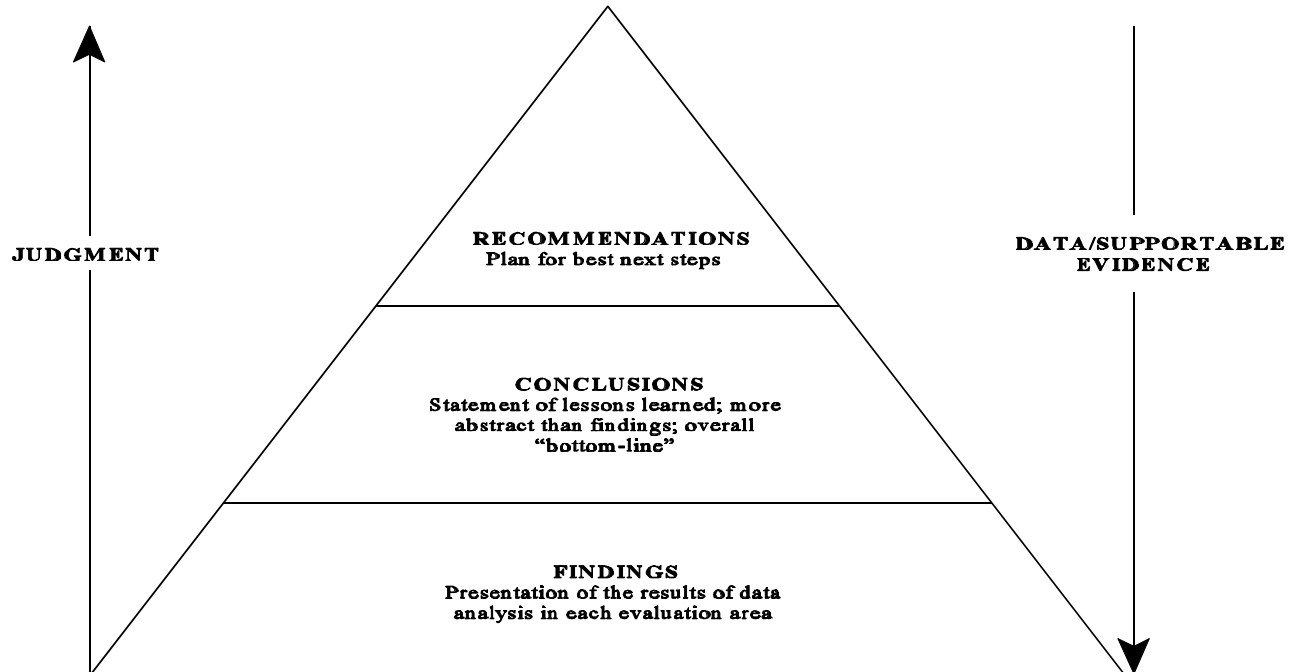
### **3. DEVELOPING CONCLUSIONS AND RECOMMENDATIONS**

Evaluation conclusions depend on more than tests of statistical significance. To conclude that a treatment initiative is better (or not) than standard treatment, or that treatment provider A is better (or not) than treatment provider B, requires that the evaluator consider and weigh a combination of qualitative and statistical evidence. Teachers of evaluation methods frequently say that statistics can neither “prove nor disprove” a theory; statistics can only lend evidence that supports or contradicts the theory. Quantitative data can lead to qualitative insights about a treatment initiative.

As evaluators seek evidence that treatment is better under one condition than another, they will need to reject numerous possible threats to validity of the comparison. Did the experimental design support conclusions that one treatment is better than another? Were data completed on a representative percentage of respondents, and were the measures themselves valid and reliable?

Ultimately, the evaluation team will need to draw conclusions and recommendations for each of the substance abuse service conditions evaluated. Conclusions and recommendations generally emerge from the analyzed qualitative and quantitative data but also incorporate the judgment and insight of the evaluation team. The relationships among findings, conclusions, and recommendations are shown graphically in Exhibit II-3. The pyramid illustrates that recommendations rest firmly on conclusions that, in turn, are solidly grounded on the findings that emerge from analyzed data. Each represents a succeeding level of abstraction and requires increasing degrees of professional judgment.

**EXHIBIT II-3  
INTERRELATIONSHIP OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS**



Finally, it should be pointed out that the modern evaluator needs to wear at least three hats when analyzing substance abuse treatment evaluation data: clinical expert, organizational psychologist, and cost-efficiency expert.

- **Clinical expert.** Some analysis questions require that the evaluator understand the clinical environment and context: how do clients flow into the treatment unit, what are the major steps in their treatment and recovery process, how are clients released from care, and what happens with clients after treatment and prior to follow-up? Understanding clinical objectives is imperative for understanding service design, delivery, and outcomes.
- **Organizational psychologist.** Other analysis questions require the evaluator to consider how individual staff view and interact within their organizational environment, and how the organization views and interacts with its broader community environment. Understanding an organization’s culture will help the evaluator understand *why* a service was implemented the way that it was, or to understand factors that differentiate organizations that may partly account for differences in client outcomes.

- **Cost-Efficiency Expert.** The third hat is analyzing cost data on both the cost of treatment services and the economic value associated with certain substance abuse treatment outcomes.

These analytic “hats” go along with the requirement that evaluators bring an understanding of survey methodology, ethnography, statistical analysis, and policy analysis to bear when summarizing the results of an evaluation. Because there is such a variety of disciplines embodied in substance abuse treatment evaluation, it is not surprising to find that some evaluators concentrate more or less on different aspects—or they wear one hat more than the others. In the words of evaluation guru Michael Scriven (1993), “Subject matter expertise may be the right hand of education programs and proposal evaluation, but one cannot wrap things up with a single hand.” Hence, an evaluation team may be a better approach than a single evaluator approach.

The following chapter is designed to provide a cross-discipline summary of evaluation questions and analytic approaches that should help evaluators get started in analytic areas that are perhaps not their first thought.

### **III. SPECIFIC QUESTIONS AND ANALYTIC APPROACHES**

The purpose of this chapter is to demonstrate specific approaches to the analysis of evaluation data so as to address the evaluation questions. The chapter is organized according to the analytic framework described in Chapter II and includes the following sections:

- Describing the planned service
- Assessing service delivery
- Assessing the impact of services
- Applying good judgment in the use of data.

Each section contains a discussion of relevant analytic principles followed by analysis planning “templates.” The templates are designed to assist the evaluator in planning or executing analyses of evaluation information. Each template contains five discrete sections that describe: (1) why the analysis of the evaluation data to address the evaluation question is important; (2) analytic questions that further amplify the evaluation question; (3) IEM tools that will assist with the analysis; (4) information sources that will provide the evaluation data for the analysis; and (5) analytic approaches.

The last section of this chapter, applying good judgment in the use of data, discusses factors that influence the rigor of the data, such as the evaluation design and the validity and reliability of the data collected. These are critical issues that should be at the top of the evaluation planning agenda. This document focuses on using data that have been collected and on assuring that measurements will support the types of analyses to be performed later, and hence the discussion of evaluation design and measurement quality are framed here for post-hoc analysis.

#### **1. DESCRIBING THE PLANNED SERVICE**

A report on the evaluation of knowledge-generating substance abuse treatment services needs to communicate to readers the design and objectives underlying the services that were evaluated. Rossi and Freeman (1993) point out that:

To devise an ameliorative social program, and to assess whether a program is working, it is first necessary to define the problem precisely, assess its extent, describe the elements of the intervention, and accurately define the target population to be served.

What were the original expectations for services and outcomes, against which actual results would be compared? The design and objectives include forward-looking statements as well as hypotheses about the expected types of service effects. More specifically, these forward-looking statements should clearly summarize how one treatment group or service is believed to be different or better than another service.

This descriptive information about planned services is necessary both for an evaluation of a new service initiative as well as for comparison of existing initiatives. For new services or treatment initiatives, the evaluator should describe the underlying rationale for the initiative and identify important steps in its implementation. The evaluator needs to describe the stages of development, specify the goals, predict relationships between services and impacts, and describe the plan by which the services were to be implemented. For comparisons of existing treatment services, the focus shifts from how the services would be implemented to how and why services are expected to be different between treatment groups.

In obtaining information for this component of the evaluation, evaluators often have to choose a primary source of information, such as proposals, interviews with staff from funding agencies, and early planning documents. Hindsight is often necessary to pull out aspects of how the services were designed. Moreover, evidence of evolution in plans or changes in circumstances need to be assembled, if the evaluator is to draw inferences about how these changes influenced implementation and outcomes.

The evaluator who is attempting to describe the planned services needs to critically analyze the “theory” and underlying assumptions of a treatment service or new initiative (Chen, 1990a and 1990b). What are the problems that the intervention will fix, and how would one know if the service was successful? Evaluation writers recommend several approaches that are more similar than dissimilar, including the logic model and the impact model. Logic modeling is the construction, either graphically or in text, of a flow chart that identifies and then describes how inputs, processes, outputs, and intermediate and longer term outcomes are related.

Using different language but describing a similar thought process, Rossi and Freeman (1993) recommend that evaluators prepare an “impact model” that articulates causal hypotheses, intervention hypotheses, and action hypotheses. Causal hypotheses state how a specified problem came into being. Intervention hypotheses state how a service intervention will affect

determinants in the underlying causal hypothesis. Action hypotheses state how the intervention will change the outcome.

Logic modeling should be incorporated into evaluation analysis. Accordingly, question 1.3 in Exhibit III-1 provides suggestions on how to describe the underlying logic model underlying substance abuse treatment services.

When presented in a report, summaries of planned services typically go hand in hand with summaries of the actual delivery of services. For example, the evaluator may wish to prepare a series of tables or matrices to assist in contrasting original plans with actual services. This includes not only quantitative results of services, but also descriptions of how, when, or by whom certain actions would be brought about. Picking out a handful of critical factors in service design and implementation and discussing what actually transpired will be invaluable to explaining outcome data later.

To help the evaluator prepare to answer questions about the planned services, however, these questions are separated from subsequent questions on how the services were implemented. Questions critical to describing the plans include:

- How was the service developed?
- What are the goals and objectives of the service?
- What was the predicted relationship between the service and expected outcomes?
- What was the plan for implementing and delivering the services?

An analytic approach to addressing each of these evaluation questions is presented in the following pages.

<b>EXHIBIT III-1 ANALYTIC APPROACH</b>	
<b>1.1 HOW WAS THE KNOWLEDGE-GENERATING SERVICE DEVELOPED?</b>	
<b>Why Important</b>	<p>Development of the knowledge-generating treatment service refers to the original planning and needs assessment that likely occurred before an organization submitted a proposal for funding. In many cases, the design of the service continued evolving as leaders continued thinking through how to make the initiative work.</p> <p>The evaluator needs to convey enough of the development process so that readers can understand the circumstances that originally led to the need for the service, how the organization responded, and why. Moreover, for initiatives that turned out considerably different from originally planned, evaluators may find useful insights in hindsight regarding the planning process that would help other organizations.</p>
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What needs assessment activities were conducted prior to beginning the knowledge-generating treatment service?</li> <li>■ What activities were involved in the development process?</li> <li>■ Which staff members participated in the development process?</li> <li>■ What problems were experienced during development and how were they resolved?</li> <li>■ What key decisions were required by the development process?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Memoranda or testimonials from internal leaders, staff, or external stakeholders</li> <li>■ Interviews with treatment services management</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Begin with simple description of steps, information sources, and decisions that were made.</li> <li>■ Consider whether there were views and perspectives that were not fully incorporated. Were external market or stakeholder concerns considered part of the planning process?</li> <li>■ Consider how prior experiences in the SDU played into the conceptualization and design of the services.</li> <li>■ Consider whether and how SDUs went outside their normal environment to obtain information regarding any new service initiatives. Did they turn to research literature, other treatment providers, or consultants for assistance?</li> </ul>

<b>EXHIBIT III-1 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>1.2 WHAT WERE THE GOALS AND OBJECTIVES OF THE KNOWLEDGE-GENERATING TREATMENT SERVICE?</b>	
<b>Why Important</b>	The first step in gauging the relative success of one treatment approach versus another is to communicate to the reader what different groups or treatment conditions were designed to accomplish. This is the first standard by which any treatment service should be judged.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What service gaps were identified that indicated a need for the knowledge-generating service?</li> <li>■ How much of the knowledge-generating services existed prior to the evaluation time period, and what change is expected as a result of the treatment service?</li> <li>■ What were the specific activities or interventions of this knowledge-generating treatment service?</li> <li>■ If the knowledge-generating treatment service is a modification of a previous substance abuse treatment service, what were the similarities and/or differences?</li> <li>■ What were the goals in terms of treatment service activities that will change?</li> <li>■ What were the goals in terms of improvements in client or administrative outcomes?</li> <li>■ Over what period were changes expected to take place?</li> <li>■ Who were the target population(s) and how are services specially designed to meet their needs?</li> <li>■ What efforts were made to ensure cultural appropriateness?</li> <li>■ Were treatment services available for children of substance abusers? Collaterals?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Memoranda</li> <li>■ Treatment service plans</li> <li>■ Initial CSAT proposal and responses to technical questions</li> <li>■ Interviews with treatment service leaders</li> </ul>

**EXHIBIT III-1 (CONTINUED)**  
**ANALYTIC APPROACH**

**1.2 WHAT WERE THE GOALS AND OBJECTIVES OF THE KNOWLEDGE-GENERATING TREATMENT SERVICE? (CONTINUED)**

**Analytic Approaches**

- Begin with a simple recitation of the goals and objectives as originally stated, then go on to describe modifications to the goals and objectives, pointing out reasons or lessons that were learned along the way.
- If treatment services did not have concrete goals, or the specific types of changes that were expected as a result of the initiative were never articulated, work with the treatment provider to determine what the original goals would have been.
- Document modifications made in the way that the SDU operates.
- Articulate what the initiative was expected to produce in terms of improved outcomes or greater efficiencies.
- Consider whether the goals were realistic?
- Consider whether the success of this initiative linked to other strategic goals of the SDU in ways that put added pressure on achieving success.
- Describe whether and how special consideration was given to cultural competency and incorporation of family or significant others in treatment. (Both of these factors may play a specific role in the treatment initiative, or they may be changing at the treatment provider level in ways that influence the treatment initiative. For some providers, these factors may not play a role at all.)

<b>EXHIBIT III-1 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>1.3 WHAT WAS THE PREDICTED RELATIONSHIP BETWEEN THE KNOWLEDGE-GENERATING TREATMENT SERVICE AND THE EXPECTED OUTCOMES?</b>	
<b>Why Important</b>	Understanding predicted relationships is critical for designing and interpreting the results of an evaluation. Understanding theoretical or predicted relationships allows the evaluator to identify and separate intended effects from unintended effects or byproducts of a service. Evaluators should communicate how the specific elements of the treatment service initiative are supposed to produce specific outcomes.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What underlying theory supports the treatment design for the target population?</li> <li>■ How are the treatment services related to the treatment service goals and objectives?</li> <li>■ How will the treatment service activities produce the expected outcomes?</li> <li>■ What evidence of previous success or lack of success is available?</li> <li>■ Has a logic model been produced that represents the linkage between the characteristics of the knowledge-generating substance abuse treatment service and the expected outcomes?</li> <li>■ What expected outcome measures have been developed for each of the treatment service components?</li> <li>■ Based on the logic model, what changes are predicted for:               <ul style="list-style-type: none"> <li>– Completion, participation in treatment, retention, and length of stay</li> <li>– Substance abuse levels and attitudes</li> <li>– Physical health</li> <li>– Mental health</li> <li>– Social relationship</li> <li>– Criminal justice involvement</li> <li>– Costs and efficiency</li> </ul> </li> </ul>
<b>IEM Tools</b>	<p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i></p> <p><i>Using Logic Models in Substance Abuse Treatment Evaluations</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Treatment service plans, proposals, background literature—or searching background literature if not already done</li> <li>■ Interviews with treatment service managers</li> </ul>

**EXHIBIT III-1 (CONTINUED)**  
**ANALYTIC APPROACH**

**1.3 WHAT WAS THE PREDICTED RELATIONSHIP BETWEEN THE KNOWLEDGE-GENERATING TREATMENT SERVICE AND THE EXPECTED OUTCOMES? (CONTINUED)**

**Analytic Approaches**

- Map (graphically or in text) how specific components of the service will lead to specific outputs and specific outcomes. This is a very conceptual analytic exercise.
- Consider what evidence will demonstrate “success” at each stage. How would the evaluator know if the component was implemented? How would the evaluator know if that component led to a specific output and a subsequent outcome?  
  
Example: Project New Beginnings hoped that a new family therapy initiative would improve outcomes by improving the recovery environment. A logic map would list the inputs and treatment service activity changes that were necessary (e.g., staff hired or qualified staff person assigned to that function; clients matched with that person); outputs that were expected (e.g., more family therapy sessions would be held than before); and outcomes that were expected (e.g., reduced problems with family members, more stable recovery environment, and better post-treatment alcohol and drug use rates compared to before implementation).
- Develop hypotheses around amounts of change in outcomes that should occur. What does the literature or other experience lead one to expect?
- Consider the intended effects. This information should ultimately be used in analysis of outcome data to determine whether the services had impacts that surpassed narrow expectations and why.
- Think through expected results to help identify and diagnose problems. Understanding what is expected of an initiative helps monitor whether expectations are being met, and if not, trace back to the underlying service components.
- Predict outcome rates (e.g., change in percent of clients using drugs and alcohol post-treatment) to help think through research design, sampling power analysis, and measurement approaches.

<b>EXHIBIT III-1 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>1.4 WHAT WAS THE PLAN FOR IMPLEMENTING AND DELIVERING THE KNOWLEDGE-GENERATING SERVICE?</b>	
<b>Why Important</b>	Treatment service providers interested in implementing new initiatives would benefit from tapping the experience of how other SDUs anticipated roll-outs. Moreover, for the evaluation, this information is necessary to help readers compare similar treatment services. Evaluators can use this as part of a process evaluation to determine whether the intended changes in services actually transpired.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What treatment services were to be offered and in what quantities? Over what time period, including start and end dates?</li> <li>■ What was the organizational structure for the treatment services?</li> <li>■ What was the planned length of stay?</li> <li>■ How were clients seeking entry to be managed? Waiting lists? Interim services? Screening?</li> <li>■ What are the characteristics of (a) intake, (b) screening, (c) case management, (d) continuing care, (e) transitional care components?</li> <li>■ How were treatment planning and discharge planning to be conducted?</li> <li>■ How were other services to be coordinated, including primary health, mental health, and social services?</li> <li>■ What is the planned continuum of treatment services?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Treatment service plans, proposals, background literature—or searching background literature if not already done</li> <li>■ Interviews with SDU leaders</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Begin with simple descriptions of the planned differences between groups, and in the case of new initiatives, how those plans would be rolled out to implementation.</li> <li>■ Profile what a typical service encounter or episode would include across different groups.</li> <li>■ Consider the duration of roll-out, timing with other changes or impacts on treatment service providers, and the appropriateness of planned management oversight and staff training given the task to be performed.</li> </ul>

## **2. ASSESSING SERVICE DELIVERY**

This section poses the question: How have the services truly been implemented? The prior section focused on identifying the goals and plans, particularly focusing on how groups or treatment conditions were expected to be different. This section seeks information regarding the veracity of the differences in services between groups. For many evaluators, this question can be cast as part of the “process evaluation.” This part of the analysis should determine:

- Whether the services reached the defined population
- Whether the services were consistent with design specifications
- What resources were expended (Rossi & Freeman, 1993).

Assessments of service delivery are necessary for evaluations of new initiatives as well as comparisons of existing services. For new initiatives, evaluators must describe how services actually changed over a baseline condition. For comparisons of existing services, evaluators must determine how many of what types of services were provided to clients in different groups. Ultimately, for either type of evaluation, the evaluator will use this information to conclude whether expected differences in services between groups and conditions actually occurred.

Evaluators should use the results of such service assessments in the analysis of outcome data later. To illustrate, if expected differences in services between groups are not observed, then any detectable differences in outcomes between groups is spurious. Conversely, if there are differences in services but there are not differences in outcomes, then the evaluator can conclude more confidently that the services did not produce the desired effect. If both services and outcomes are different, then the evaluator can use service information to draw out an explanation of the impacts. By compiling variables in a manner that allows for dose-response style analyses, the evaluator can assess whether clients who received more fared better in treatment.

Through the assessment of service delivery, the evaluator will basically pronounce if there is sufficient reason to believe that the services between groups being compared are truly different. Such a pronouncement requires that the evaluator present a convincing case to reject threats to the validity of the comparison. Rossi and Freeman (1993) identified at least three types of “delivery system errors” that could render outcome comparisons meaningless: non-services, wrong treatment, and unstandardized treatment; we add a fourth, premature treatment:

- In non-services, money was spent but no one can articulate on what it was spent or what services changed as a result
- With “wrong” treatment, staff implemented a different model than originally specified or staff implemented an inappropriate treatment
- With unstandardized treatment, staff implemented widely divergent treatments. Some delivered treatment according to specification, others did not
- In premature treatment, data are collected on treatment effects before a new treatment service has matured and the kinks worked out.

Along the way to rejecting the possibility of a delivery system error, the evaluator will need to consider the range of internal and external forces that resulted in the changes or delayed certain activities in ways that may have undermined the services. Often, the lessons of how and why this happened produce useful insights for future program and evaluation planning.

One framework for thinking about the types of analysis to be performed is to separate the analyses of inputs, activities, and outputs. Inputs include the flow of clients and resources into a treatment service; the activities are micro processes (e.g., client/clinician interactions), and outputs are the numbers of services provided to a number of clients. Thinking in this framework should assist the evaluator in considering a chain of evidence. Following is an example that illustrates how to analyze the input-activity-output information.

Staff in the XYZ Agency report that services were fully implemented, yet the attendance or numbers of clients served were far less than had been expected. There is a logical disconnect that the evaluator will need to understand. The evaluator may find on further exploration that referrals were not being made into the treatment in the way that had been expected and that the nature of clients being referred had changed. So, while staff were working hard serving the clients, there had been a breakdown in the system before the clients got there in the first place.

Performing assessments of service delivery requires that quantitative data be paired with qualitative information. Observational information from the evaluator, combined with the personal experiences and observations from staff, management, clients, and stakeholders, are all potentially useful. Evaluations should also strive to collect client-level data regarding services provided, including types of services, length, and frequency. This often means going to client

records or relying on client interviews. Some treatment providers have sophisticated billing systems that allow for such information retrieval.

As noted in Chapter I, it is important that the evaluator use “triangulation” to acquire the most accurate picture using qualitative insights about service delivery. Triangulation requires multiple sources of information that essentially corroborate each other. The example of the XYZ Agency illustrated the use of triangulation on the staff perception of program effects: Client data and information about referral sources shed additional light on the staff perspective.

Assessments of service delivery also require insight regarding clients receiving services and treatment group assignment. Often, demographic data and baseline assessments of client severity are necessary to profile the populations who receive services. It is then incumbent upon the evaluator to consider how the characteristics of the clients—or how they entered their respective treatments—contributes to, or results from, differences in treatment services.

For service delivery assessment, it is important to be familiar with the following concepts that were developed by CSAT to aid in the understanding of substance treatment:

- Service delivery unit (SDU): a single level of care provided at a single site.
- Treatment episodes: typically, clients enter and exit treatment services multiple times; this recognition reshaped CSAT’s concept of treatment discharge as an end point for treatment experiences and as a starting point for follow-up data collection and the measurement of treatment outcomes and impacts.
- Treatment “bundles”: complementary to the SDU measurement concept is the fact that treatment episodes frequently involve a continuum of care from detoxification to aftercare and that treatment clients may receive a “bundle” of services from several SDUs during one treatment episode.

In assessing service delivery, the evaluator needs to answer the following questions:

- How were services implemented?
- What were the characteristics of persons who sought treatment?
- What were the characteristics of the treatment process, and how does the process compare to what was planned?

- What are the characteristics of the staff?
- What are the characteristics of the implementation agency?
- What is the external context surrounding implementation of services?
- How much did the services cost?

The following pages illustrate the analytic approach to answering each of these questions.

<b>EXHIBIT III-2 ANALYTIC APPROACH</b>	
<b>2.1 HOW WAS THE KNOWLEDGE-GENERATING SERVICE PLAN IMPLEMENTED?</b>	
<b>Why Important</b>	Understanding how SDU activities changed will inform subsequent analysis of outcomes and cost data. This information is also useful for other SDUs to understand before they implement their own similar initiatives: How did activities change in the SDU? What lessons were learned during implementation?
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What was the implementation process?</li> <li>■ What instructions, protocols were used?</li> <li>■ What training and orientations were provided at the start and during the course of the knowledge-generating service?</li> <li>■ What incentives were provided to staff, clients?</li> <li>■ What organization changes were required, e.g., staff assignments, reporting, communications?</li> <li>■ What major changes and adjustments in treatment services were required?</li> <li>■ What factors facilitated and what factors impeded effective implementation?</li> <li>■ What was not implemented and why?</li> <li>■ When did services first change? When were clients first affected?</li> <li>■ How long did the knowledge-generating service last?</li> <li>■ Did the knowledge-generating service alter course over its life?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ SDU records</li> <li>■ Interviews with SDU leaders</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Begin with simple descriptive information that answers the questions.</li> <li>■ For change-based evaluations, prepare a chronology of events that summarizes milestones and decisions and important staffing or other contextual changes.</li> <li>■ For evaluations that compare groups, consider whether surprises that occur during the study are important for understanding results.</li> <li>■ Consider factors that helped foster success or that were barriers to success, and describe how the SDU staff either capitalized on the positive factors or dealt with the barriers.</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>2.2 WHAT WERE THE CHARACTERISTICS OF THE INDIVIDUALS WHO SOUGHT TREATMENT?</b>	
<b>Why Important</b>	The specific methods by which clients were recruited into the SDU are critical for establishing credibility of most client-oriented evaluations. Moreover, in initiatives where reaching a specific target population is part of the initiative, these data constitute an important measure of success. Answering these questions also will help the evaluator demonstrate how their treatment population is different from others or how the clients in the knowledge-generating initiative differ from other clients.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did the client population match the intended target population?</li> <li>■ What was the number of clients who received the following compared to target population?                             <ul style="list-style-type: none"> <li>■ Outreach contact</li> <li>■ Intake/assessment</li> <li>■ Referral</li> <li>■ Treatment</li> </ul> </li> <li>■ Were these clients part of the intended target population?</li> <li>■ What marketing and advertising strategies were used?</li> <li>■ What outreach was provided or procedures put in place to identify and recruit clients for the knowledge-generating service?</li> <li>■ How did clients gain access? Who referred them and what criteria did they have to meet to gain access?</li> <li>■ Were there problems in the fit between knowledge-generating slots and clients who needed them?</li> <li>■ How long did clients have to wait before entering treatment? Was there a waiting list and how was it managed?</li> <li>■ How many clients were refused treatment and/or sent to another SDU? Why?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Client records</li> <li>■ MIS</li> <li>■ SDU records and notes</li> <li>■ Staff perceptions</li> </ul>

**EXHIBIT III-2 (CONTINUED)**  
**ANALYTIC APPROACH**

**2.2 WHAT WERE THE CHARACTERISTICS OF THE INDIVIDUALS WHO SOUGHT TREATMENT? (CONTINUED)**

**Analytic Approaches**

- Analyze frequencies and distributions of client data for the entire population first and then for subgroups.
- Look for comparisons and contrasts between the SDU population and research or other information about clients elsewhere (e.g., are clients in this SDU different based on demographics, substance abuse, or other factors from other treatment clients)?
- Compare target populations with other populations within the SDU, and compare treatment initiative clients with standard treatment clients.
- For substance abuse indicators, identify main problem drugs and frequency of use.
- Profile the typical client’s demographics (e.g., the average client is a 33-year-old African American unemployed cocaine addict). This provides a useful snapshot of the typical client for the reader. Then, consider the distribution of some important characteristics: If the average age is 33, is that a unimodal distribution or is there a mixture of younger clients and older clients? Even if cocaine is the drug of choice among most clients, is there enough use of other drugs to be worth noting?
- Rely on staff insights and stakeholders (e.g., referral sources) to understand how clients gained entry to the treatment service or to different treatment groups.
- Look for correlations between problems that clients have at intake, and consider how these correlated problems relate to goals of treatment. For example, if most of an SDU’s cocaine addicts are pregnant women who are also depressed, then one would predict better outcomes if mental health needs are also addressed.
- Describe the important steps by which clients get access to treatment and how factors either facilitate or impede access. Compare how knowledge-generating clients enter versus standard clients and how any recruitment into groups is conducted.
- Consider whether there is potential for bias resulting from how clients enter different groups and whether they participate in evaluation data or not.

<b>EXHIBIT III-2 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>2.3 WHAT WERE THE CHARACTERISTICS OF THE TREATMENT PROCESS AND HOW DO THEY RELATE TO WHAT WAS PLANNED?</b>	
<b>Why Important</b>	Quantifying and describing the “active ingredients” of treatment services actually received by clients is necessary for interpreting outcome data, particularly data that compare outcomes across two groups or across multiple SDUs. Without this information, it is very difficult to determine whether differences in outcome data are due to chance, survey design, or treatment effects. Alternatively, if outcomes are not different, these data are needed to verify that clients who participate in a treatment service initiative did indeed receive different services.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What types of services were provided, for how long, in what quantities?</li> <li>■ Were services commensurate with client needs as assessed at intake?</li> <li>■ What services were heavily underutilized?</li> <li>■ How did actual service delivery relate to the planned services? What changed over the baseline service delivery?</li> <li>■ Which treatment services have been implemented? Which treatment services have not been implemented?</li> <li>■ Were planned services offered in the planned quantities?</li> <li>■ What steps were taken to assure culturally, gender-, and age-appropriate services?</li> <li>■ How are substance abuse treatment rehabilitation services coupled with physical, mental, social services? What gaps exist in services?</li> <li>■ What specific services has each client received?</li> <li>■ How many individual and group sessions were held, and what proportion of possible sessions did clients actually attend?</li> <li>■ Are children or significant others involved in treatment? How?</li> <li>■ What are the characteristics of (a) intake, (b) screening, (c) case management, (d) continuing care, (e) transitional care components?</li> <li>■ How are treatment planning, treatment exit procedures conducted?</li> <li>■ How long did clients wait on average to access services?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i> <i>Minimum Evaluation Data Set: Core Data Lists (see Client level services received variables)</i>

<b>EXHIBIT III-2 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>2.3 WHAT WERE THE CHARACTERISTICS OF THE TREATMENT PROCESS AND HOW DO THEY RELATE TO WHAT WAS PLANNED? (CONTINUED)</b>	
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Client records</li> <li>■ Client interviews or focus groups</li> <li>■ SDU management information systems</li> <li>■ SDU manager interviews</li> <li>■ Staff interviews</li> <li>■ Billing records</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Summarize client-level data on the actual delivery of services, using descriptive statistics.</li> <li>■ The evaluator should only rely on “planned” services only if client-level data do not exist. For example, the statement “a typical client <i>should</i> receive 1 hour of individual and 2 hours of group counseling per week” is a <i>planned</i> service. Actual clients may not receive those services at that frequency. Some could have received more; there likely were missed appointments and other factors that made the services less meaningful.</li> <li>■ Conduct subgroup analyses on treatment groups and target populations. How many of which types of services did clients in groups A versus B receive?</li> <li>■ Consider evidence that specific types of services may have been heavily underutilized. Are there systematic patterns to support this, and what are the possible explanations?</li> <li>■ Consider services that were delivered in light of client problems: Are there clear gaps that emerged from analysis of data, corroborated by staff perceptions?</li> <li>■ Create analysis variables that summarize levels of services provided to clients. For example, if there are 15 identifiable services available, what proportion of those did the typical client receive? Or, add up the hours of individual counseling received, and consider dividing this by length of time in treatment to obtain a dose “rate.”</li> <li>■ Consider how these data compare against planned services and against the types of activities that were changed, as noted in Question 2.1? Do services data generally support or contradict that there are differences in services between groups in ways that were anticipated? Why or why not?</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>2.4 WHAT WERE THE CHARACTERISTICS OF THE STAFF?</b>	
<b>Why Important</b>	Staff who deliver the “active ingredients” of treatment bring different levels of experience and orientation to bear. Moreover, they are influenced by their environment. Their view of the SDU as a place to work, as a positive recovery environment, and their views of the environment shape their work. The manner in which they deliver services and administer SDU rules will help the evaluator understand the outcomes that were attained.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ How many staff provide treatment services? What are their educational and experiential backgrounds?</li> <li>■ Are the staff census, staff ratios, and staff backgrounds consistent with the treatment service design?</li> <li>■ What efforts have been made to improve staff development and staff quality?</li> <li>■ What are the staff hiring, turnover, and vacancy rates? What efforts have been made to improve staff recruitment and retention?</li> <li>■ How is staff morale and job satisfaction? How do staff view their organization and recovery environment?</li> <li>■ What are the primary therapeutic approaches employed by staff?</li> <li>■ How large are client caseloads?</li> <li>■ What is the balance of time spent by clinicians on direct care, management of clients, and paperwork or administrative duties?</li> <li>■ What are the dominant therapeutic approaches and clinical priorities held by clinicians; how tailored is treatment to individual needs, and what other factors illustrate how clinicians interact with clients?</li> <li>■ How do staff view the efficacy and quality of the service environment?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i> <i>Minimum Evaluation Data Set: Core Data Lists (section IV)</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ SDU manager interviews</li> <li>■ SDU staff interviews</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>2.4 WHAT WERE THE CHARACTERISTICS OF THE STAFF? (CONTINUED)</b>	
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Characterize the SDU using clinician data. “Clinicians typically employ ____ approach to treatment, bringing ___ years of experience on average to bear.” About ___ percent have a college degree and ____ are professionally credentialed. Also characterize hiring, turnover, vacancy, recruitment, and morale/satisfaction.”</li> <li>■ Describe whether staff factors have been changing, particularly in projects that introduce changes. Changes in staffing patterns are important for establishing the nature and level of continuity in the environment.</li> <li>■ Explore differences among staff across treatment groups regarding their levels of experience, training, and perceptions of the work environment.</li> <li>■ Try to characterize the work culture as it relates to individual staff factors, and differences in treatment between groups. Is there insight regarding patterns of communication, routinization of functions, and shared values among staff that can be used to describe differences in groups or explain implementation of changes?</li> <li>■ Try to characterize informal and formal leaders within the organization and how their involvement relates to the success of an initiative.</li> <li>■ Consider how turnover relates to service quality, in terms of staff training, learning curve, and continuity in general as well as with specific clients. Turnover, like many factors that indicate the staff’s perception of the work environment, may be a dependent or independent variable in analysis. Some evaluators may use turnover as a predictor of outcomes (higher turnover leads to worse outcomes); alternatively, it would be useful to consider whether outcomes predict turnover (SDUs with more “successes” are likely to retain staff).</li> <li>■ Consider how questions dealing with work environment, job satisfaction, and morale relate to one another.</li> <li>■ Consider how questions dealing with professional training and therapeutic approach relate to one another.</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>2.5 WHAT WERE THE CHARACTERISTICS OF THE IMPLEMENTATION AGENCY?</b>	
<b>Why Important</b>	<p>“Characteristics of the agency” refers to the context in which services are implemented and to the recovery environment offered to clients. Assessing these characteristics is necessary for understanding factors that account for the degree of implementation of an intervention, and for controlling for the influence of the recovery environment in explaining the differences in outcomes achieved across sites.</p>
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What are the physical attributes of the substance abuse treatment service site?</li> <li>■ What is the steering committee structure, membership, and <i>modus operandi</i>? At what developmental stage is the steering committee? What is the effectiveness of the steering committee?</li> <li>■ Under what authority is the knowledge-generating substance abuse treatment service operating?</li> <li>■ What accountability issues affect knowledge-generating treatment services? Are there specific requirements for spending/accounting for specific dollar amounts?</li> <li>■ How has the organization changed over the life of the knowledge-generating treatment service?</li> <li>■ What is the ownership and profit status of the organization?</li> <li>■ Which funding sources account for most of the organization’s revenue, and by what financing method?</li> <li>■ What are the qualifications, experience, and tenure of organization leadership?</li> <li>■ How culturally competent is the agency relative to its service area?</li> </ul>
<b>IEM tools</b>	<p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i> <i>Minimum Evaluation Data Set: Core Data Lists (section III)</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ SDU counselor interviews</li> <li>■ SDU management interviews</li> <li>■ Board and stakeholder interviews</li> <li>■ Evaluator observations</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>2.5 WHAT WERE THE CHARACTERISTICS OF THE IMPLEMENTATION AGENCY? (CONTINUED)</b>	
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Bring about narrative discussion of most analysis questions.</li> <li>■ Describe degree of change versus stability in the organization characteristics over the treatment service’s tenure, and how these changes may have affected implementation and outcomes. For example, if revenues decreased remarkably, this may have strained resources, resulting in clients in general receiving fewer or less timely services.</li> <li>■ Consider and then describe any relationship of ownership, management, and leadership structures to the treatment service design and implementation.</li> <li>■ Determine if the size of the organization and relationships with parent organizations affect the services, affect client flow and characteristics, and impact upon implementation of service changes.</li> <li>■ Express any financial information (e.g., question 2.5.7) in percentages. For example, report the proportion of revenue by source for any source that constitutes more than 10 percent of revenue (leaving the balance in “other”).</li> <li>■ Use several of the variables (e.g., funding sources, ownership, stability of leadership) as independent variables in statistical analyses with data pooled across multiple sites to control for SDU effects on implementation and outcome.</li> <li>■ For change-based evaluations, consider how information regarding these questions affected implementation of a new service.</li> </ul>

<b>EXHIBIT III-2 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>2.6 WHAT WAS THE EXTERNAL CONTEXT IN WHICH THE KNOWLEDGE-GENERATING SERVICES WERE IMPLEMENTED?</b>	
<b>Why Important</b>	SDU operations are influenced by circumstances in the external environment and how the SDU relates to other organizations. Evaluators need to be familiar enough with important developments in the regulatory, business, and service environments that may affect SDUs. They should be able to determine how those factors relate to implementation of the knowledge-generating treatment service and how outcomes could be influenced by the outside environment.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What is the service delivery (geographic?) area for the knowledge-generating substance abuse treatment service?</li> <li>■ What are the economic characteristics of the area (e.g., employment trends and unemployment rates)?</li> <li>■ What substance abuse treatment and related services exist in the area?</li> <li>■ What linkages exist between the service delivery agency, the community, and other agencies?</li> <li>■ What are the relationships between the knowledge-generating substance abuse treatment service and the services of other agencies (e.g., primary health, mental health, and other social service agencies)?</li> <li>■ What mechanisms link these services (e.g., cooperative agreements)?</li> <li>■ How many agencies are formally/informally linked?</li> <li>■ How has the quality of knowledge-generating substance abuse treatment services and other services been affected by these linkages?</li> <li>■ How many clients have been referred, and how many served as a result of these linkages?</li> <li>■ What has been the impact of these linkages on client matching?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Needs assessments</li> <li>■ Local chambers of commerce (economic conditions)</li> <li>■ Local government offices (employment statistics and economic conditions)</li> <li>■ SDU managers, staff, and management reports</li> </ul>

**EXHIBIT III-2 (CONTINUED)**  
**ANALYTIC APPROACH**

**2.6 WHAT WAS THE EXTERNAL CONTEXT IN WHICH THE KNOWLEDGE-GENERATING SERVICES WERE IMPLEMENTED? (CONTINUED)**

**Analytic Approaches**

- Describe the context in enough detail to convey critical aspects of the environment enough to subsequently explain how the environment relates to services and outcomes.
- Consider how features of the environment play into successes or failures of treatment initiatives. Changing policies, regulations, financing, availability of substance abuse treatment, or other types of services in the market will all affect how this SDU operates, and how new initiatives will work.
- Describe how the economy and other macro issues such as availability of drugs or increases in community policing bear on client outcomes. A poor economy may make it more difficult for clients to attain employment; more police on the street may lead to more arrests for crimes.
- Consider how the SDU is linked or coordinated with other service providers. Do service linkages influence implementation of the knowledge-generating initiative, or vice versa?

<b>EXHIBIT III-2 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>2.7 HOW MUCH DOES THE KNOWLEDGE-GENERATING SERVICE COST?</b>	
<b>Why Important</b>	Evaluators should be able to articulate the cost of a service from a variety of perspectives: full economic cost of an episode, per diem costs, revenue per client, and unit costs for specific treatment components. This information should help readers understand whether and how the treatment service initiatives are more or less expensive than standard treatment. Finally, cost data are required in conjunction with outcome data to perform any cost-effectiveness (relative efficiency) or cost-benefit (cost-offset) analyses.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What are the total operating costs for the facility (all SDUs)?</li> <li>■ What is the percentage of costs attributable to administration, facility maintenance and operation, and direct client care?</li> <li>■ What is the percentage of personnel service costs to total costs?</li> <li>■ What is the ratio of clients to full-time equivalent (FTE) direct care staff and total staff?</li> <li>■ What is the average salary cost per employee classification?</li> <li>■ What is the average SDU cost per client, per episode, and per day in treatment?</li> <li>■ What is the cost of service per treatment component, including intake, treatment, and aftercare?</li> <li>■ What is the cost of service by unit of service (per SDU and client)?</li> <li>■ Is the budget sufficient to operate treatment services as designed?</li> <li>■ How do SDU costs for knowledge-generating activities compare to costs for standard treatment?</li> <li>■ How financially stable is the SDU, e.g. months of available resources and balance of revenues and expenditures?</li> </ul>
<b>IEM Tools</b>	<i>Substance Abuse Treatment Cost Analysis and Allocation Template</i> <i>Minimum Evaluation Data Set: Core Data Lists (section V)</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Financial records for the SDU</li> <li>■ Financial records for the parent organization</li> </ul>

**EXHIBIT III-2 (CONTINUED)**  
**ANALYTIC APPROACH**

**2.7 HOW MUCH DOES THE KNOWLEDGE-GENERATING SERVICE COST? (CONTINUED)**

**Analytic Approaches**

- Compile data over a fiscal year if possible.
- Allocate costs between treatment conditions, applying values that are entirely allocable and then dividing remaining costs based on percent of clients seen.
- If this is a change-based evaluation, compile a baseline cost estimate and then a subsequent cost estimate to determine if costs changed. Analyze whether other factors also changed that could have altered the cost profile, e.g., change in salary levels, increases or decreases in numbers of staff.
- Prepare several ratios for analysis across groups.
  - Calculate total cost, and divide this by number of clients served during the year to obtain cost per client.
  - Calculate the number of FTE direct care and total staff, and divide each by the number of clients currently enrolled to obtain a staff to client ratio.
  - Add allocated salaries, and divide by the number of allocated FTEs to determine average pay per FTE.
  - Add allocated salaries and fringe for each group, and divide by clients served to obtain labor cost per client served.
  - Calculate administrative costs and divide by total cost to obtain a ratio of “administrative efficiency.”
  - Divide total cost by direct services costs to obtain a “fiscal multiplier.” Aside from comparing across treatment services, this multiplier can be used by SDUs to set prices per unit of service (e.g., by multiplying an “unloaded” service cost by the multiplier to assure that the price charged would cover total SDU expenses.)
- Compare ratios across groups and over the life of the treatment service.
- Consider whether resources were adequate to implement the service, and whether the SDU and purchasers might consider the service to be worth any extra cost.

### **3. DESCRIBING IMPACTS**

The goal underlying outcome analyses is to determine the net effects of the services, uncontaminated by the influence of other processes and events (Rossi & Freeman, 1993). Fortunately for the substance abuse treatment evaluator, there is a rich history of outcome analysis in the field. Readers are advised to turn to several major national studies for good discussions of client behaviors and status at entry to treatment, and how those change during the course of treatment and in the years thereafter (e.g., Hubbard et al., 1989; Gerstein and Harwood, 1991, Gerstein et al., 1994).

In conceptualizing outcomes, evaluators should separately consider treatment completion status, reduction in substance abuse, and amelioration of other social impacts of substance abuse (e.g., crime, employment, health, and family and social problems). Treatment completion status—and associated measures of treatment participation, length of stay, and retention—are critical measures used by funding bodies and regulators as economical “early warning signs” of problems within specific treatment agencies, but they are also confounded by numerous systemic, exogenous influences. Reduction or elimination of substance abuse is obviously the goal of most treatment services. The third category—other social impacts, such as crime or reliance on welfare—are the problems that policy makers often seem most concerned with.

There are three major steps in the process of analyzing outcome data: descriptive analysis, service analysis, and cost analysis.

- First, calculate and describe the outcomes achieved. The outcome analyst, before delving into further analysis about factors that account for differences in outcomes, should characterize outcomes across an entire population. What proportion of clients did not use drugs at follow-up? What was the percent reduction in drug use days? Next, describe how these outcomes vary across treatment groups, and how they vary by client sub-population (e.g., men versus women).
- Second, explore hypotheses about the relationship between the services, clinicians, and service environment and the outcomes achieved. Many analysts focus on the relationship between length of stay and outcomes; there are many other clinician factors (such as their experience, education, therapeutic approach) and the service delivery unit (such as its size, approach to stages/phases of treatment) that may help explain differences in outcomes across treatment services. Experimenting with alternative combinations of variables that describe the dynamic between the client, counselor, SDU, and the SDU’s environment will help explain how outcomes vary across services.

Many service-level data are included in the IEM *Minimum Evaluation Data Set* (MEDS) particularly the clinician and SDU components of the MEDS. The analytic approach to question 3.1, below, will help the reader understand how to begin this analysis; however, this type of analysis should be performed within each of the different “domains” (e.g., family and social relationships, employment) presented in questions 3.2 through 3.7, below.

- Third, compare the costs and outcomes of treatment services (see the analytic approach to question 3.8, below). Several substance abuse treatment outcomes studies performed in recent years have included comparisons of the economic cost of treatment with the outcomes achieved, and some studies have also estimated the economic value of reduced social impacts of substance abuse. It is important for the analyst to perform this type of analysis. Two types of analysis are highlighted here: cost-effectiveness and cost-offset (or cost-benefit). A more complete discussion of these can be found in a separate IEM paper entitled *Adding “Value” to CSAT Demonstration: The What, How, and Why of Cost Analysis*.
  - *Cost-effectiveness (or cost-efficiency)* compares the relative costs with outcomes achieved across two or more types of service, or between a standard and experimental condition. The product of cost-effectiveness analysis is a ratio of outcomes to cost for each treatment service condition. Such a ratio indicates, for a single outcome, which service was a better “value.” The bigger the outcome to cost ratio, the better value. For example, an evaluator would divide the calculate the reduced drug-using days from pre- to post-treatment and divide by the cost of services. When this ratio is obtained from multiple service conditions, they can be compared: the higher ratio (more outcomes per cost) is the better value.
  - *Cost-offset (or cost-benefit)* analyses completed in this field nearly exclusively estimate the monetary value of certain social impacts and estimate the savings that accrue when these social impacts are lessened. When treatment offsets certain social impacts, benefits accrue to society, and these economic benefits can be directly compared to the dollar investment in treatment services (e.g., Gerstein et al., 1994). While this method certainly can also be used to compare services, it (unlike cost-effectiveness analysis) is also meaningful to articulate the economic benefits of a specific intervention.

It should be noted that reliable outcome data that reach valid conclusions about the impact of treatment services are required before one should attempt to relate services and outcomes or cost and outcomes. The quality and depth of cost data available (see Chapter II and the separate IEM product, *Substance Abuse Treatment Cost Analysis and Allocation Template*) will also enrich the economic analysis of outcomes.

Several general recommendations can be drawn to assist in carrying out outcome analyses. It is critical to note that several of these recommendations have “up front” implications for the design of a treatment service evaluation, particularly the design of measurement instruments.

- **Compare status or behavior across time periods, such as pre-treatment and post-treatment, making certain that reference periods are of the same duration.** If not, then the data need to be statistically adjusted to make them comparable. Because this may introduce biases based on assumptions used to correct the data, it is better to avoid this by having the time periods comparable.
- **Validate self-reported data against other objective data sources, e.g., physical specimens or official records.** Because of the biases in self-reported data, it is usually advisable to compare these responses or a portion of these responses, whenever possible to other data sources. Through this, the evaluator is often able to assess the quality of self-reported data. It is important to keep in mind that all data sources have biases and limitations.
- **Perform intra-group as well as intergroup comparisons of client status and behavior, identifying areas where clients are homogenous or heterogenous.** For similar treatment interventions with similar target populations, these comparisons can highlight possible important differences in client characteristics or service effects.
- **Clients need to be compared against their own baseline data (usually the pre-treatment period), rather than combing the data across clients within a treatment condition and then comparing them to the aggregate baseline data.** Aggregating the data obscures how many clients used less of a substance, used more, or stayed the same. Consider using statistical tests that compare differences in scores for each client.

Results from the statistical test or tests cannot be taken at face value. “Statistical significance is the easy part of establishing significance” (Scriven, 1993). The evaluator needs to examine the results. If the evaluator finds a positive treatment effect for a particular outcome, he or she has to consider whether there were any threats to the validity of the evaluation (addressed more fully in the next section). For example, given the evaluation design as executed, is there any other plausible cause that could account for the outcome? Next, the evaluator needs to determine, if the evaluation design permits, what elements of the treatment service may have produced the effect. A good process evaluation is helpful in exploring this. Another question the evaluator needs to ask is “Did any treatment subgroup do better or worse than the group as whole?” If there are

differences among subgroups, the evaluator needs to explore the possible causes for this. Finally, if the results appear to be valid, the evaluator needs to explore the implications of the findings both for the particular treatment service and the field as whole.

If the evaluator finds no treatment effect or negative treatment effect for a particular outcome, similar questions apply: “Were there any threats to the validity of the evaluation that invalidates the study or the analysis of this outcome variable?” Subgroups in the treatment population need to be examined. Did any subgroup show positive results? Did any subgroup do worse relative to others? If the results of the analysis seem valid, then the evaluator needs to explore the implications of the findings both for the particular treatment service and the field as a whole.

This outcome analysis is guided by the following questions:

- How long did clients stay, and how much did they participate in treatment?
- Were predicted substance abuse outcomes observed?
- Were predicted social and family relationship outcomes observed?
- Were predicted effects on employment and/or education observed?
- What predicted physical health effects were obtained?
- Were predicted effects on mental health observed?
- What were the effects on criminal justice involvement?
- What is the relationship between the costs and the outcomes of the knowledge-generating substance abuse treatment service?

The analysis approach for each of these questions is presented in the following pages.

<b>EXHIBIT III-3 ANALYTIC APPROACH</b>	
<b>3.1 HOW LONG DID CLIENTS STAY, AND HOW MUCH DID THEY PARTICIPATE IN TREATMENT?</b>	
<b>Why Important</b>	Treatment completion has repeatedly been shown to be related to post-treatment success, and length of stay has been shown to be related to post-treatment success for clients who did not complete treatment. While many treatment clients do not complete treatment, they do participate enough to achieve some effect. It is necessary to demonstrate participation and completion for funding agencies and regulatory bodies, and length of time in treatment may serve as a proxy if additional outcome measures are not available.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What number and percent of clients who approached the SDU:                             <ul style="list-style-type: none"> <li>– Were referred elsewhere for treatment?</li> <li>– Entered treatment?</li> <li>– Stayed in treatment 1 week, 1 month, 3 months, and 6 months?</li> <li>– Completed treatment?</li> </ul> </li> <li>■ How does treatment entry and completion vary by target population?</li> <li>■ What was the average length of stay for completers and non-completers, and how does this vary by target population?</li> <li>■ What is the ratio of average length of stay to planned length of stay?</li> <li>■ What percentage of clients participated in aftercare or other treatment follow-up services?</li> <li>■ What client and SDU factors are associated with treatment completion and longer length of stay?</li> <li>■ Of those clients who did not complete treatment, what proportion left voluntarily against medical advice (no-show, dropout)?</li> <li>■ Of those who did not complete treatment, why did they leave?</li> <li>■ Of those clients who did not complete treatment, what proportion left involuntarily (administrative discharge for drug use, rule violations, violence, etc.)</li> </ul>

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.1 HOW LONG DID CLIENTS STAY, AND HOW MUCH DID THEY PARTICIPATE IN TREATMENT? (CONTINUED)</b>	
<b>IEM tools</b>	<p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i></p> <p><i>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations</i></p> <p><i>Minimum Evaluation Data Set: Core Data Lists (section IV: Client Data)</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ SDU managers and staff</li> <li>■ SDU client records</li> </ul>
<b>Analytic Issues</b>	<ul style="list-style-type: none"> <li>■ The evaluator should document the number of clients who completed treatment and the number who do not. In addition, the evaluator should summarize the length of stay for clients who remained to completion and the length of stay for those who quit early.</li> <li>■ Many treatment services have different phases or steps in treatment, so care needs to be taken that the complete treatment episode is described. For example, a residential SDU may have a re-entry component that involves clients living at a SDU facility but working or attending school in the community, these clients may live in a separate facility from the main residential facility. These clients, however, have not completed treatment until they finish this phase of treatment.</li> <li>■ In some outpatient SDUs there may be a difference between the date a client is officially discharged from the SDU and the last date he or she attended the SDU. More generally, variation in policies on discharge for relapse and in definitions for beginning and dropping out of the SDU can complicate cross-service comparisons</li> <li>■ The reasons given for a termination prior to completion, wherever possible, should be considered in the analysis. Some reasons for termination, e.g., conviction and incarceration for a crime committed prior to treatment, are neutral. Others like discharge for not complying with SDU rules or leaving against medical advice (dropout) are negative, while referral to a less intensive modality (“stepping down”) is positive.</li> </ul>

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.2 WERE PREDICTED SUBSTANCE ABUSE OUTCOMES OBSERVED?</b>	
<b>Why Important</b>	Most substance abuse treatment interventions have as a primary goal the elimination of drug (including alcohol) consumption by their clients; all treatment services seek substantial reductions in use combined with skills for avoiding relapse.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What proportion of substance abuse treatment service clients significantly reduced their use of alcohol and/or illicit drugs at follow-up as compared to pre-treatment use? increased use? control/comparison groups?</li> <li>■ How much did clients reduce [increase] days, frequency, or quantity of alcohol and/or illicit drug consumption at follow-up relative to before treatment?</li> <li>■ How do changes in use of primary problem drug/alcohol relate to use of other drugs/alcohol? Is there a reduction in primary drug but not in other substances? Is there a substitution of one primary problem drug with another?</li> <li>■ What proportion of those recruited were free of alcohol and/or drugs following treatment? What proportion of those who began treatment? Completed treatment?</li> <li>■ Does participation in treatment services have a positive impact on clients' attitudes toward alcohol/drug use compared to their attitudes at pre-treatment? as compared to control/comparison groups?</li> </ul>
<b>IEM tools</b>	<i>Minimum Evaluation Data Set: Core Data Lists</i> (section IV: Client Data)
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Self-reported data regarding the use of alcohol and other drugs.</li> <li>■ Physical specimens (urine or hair analysis).</li> <li>■ Observations (interviewers and clinicians can observe the respondent and note if he or she appears to be "high" or if the respondent smells of alcohol).</li> <li>■ Coroner reports (for clients that died during the follow-up period to determine if this was a drug-related death).</li> <li>■ SDU records (for information concerning the consumption of drugs during treatment for outpatient SDUs, e.g., urine analysis reports, client statements to staff).</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.2 WERE PREDICTED SUBSTANCE ABUSE OUTCOMES OBSERVED? (CONTINUED)**

**Analytic Approaches**

- The consistent positive association between time in treatment and subsequent outcomes is at least partly due to third variables that cause both (e.g., client motivation). In principle, regression models can be used to examine the impact of other causes of outcome (e.g., client characteristics), but can only address causes that are measured. Consequently, special care should be taken in interpreting retention-outcome associations.
- The evaluator should attempt to characterize “participation” in treatment as a proxy outcome, much like length of stay is a proxy for other outcomes. Greater levels of participation (e.g., attendance and engagement) are markers of success, and the evaluator should determine whether participation predicts subsequent outcomes.
- Logistic regression is a useful statistical technique to analyze the impact of various variables (e.g., severity of drug use, ethnicity and age) on SDU completion. If data from multiple SDUs are being pooled, then a hierarchical modal can be used. If two services or SDUs are being compared, odds ratios can be estimated and interpreted as relative service effects after controlling for the influence of pre-treatment differences between groups on non-SDU factors.
- Survival analysis is a useful statistical technique to analyze the impact of various variables (e.g., severity of drug use, ethnicity and age) on clients’ length of stay, provided that the intended length of stay is either the same for all clients (fixed-length SDUs) or indefinite (e.g., methadone maintenance). If intended length of stay is variable, the technique is not applicable. If data from multiple SDUs are being pooled, then a hierarchical modal can be used. If multiple SDUs are being compared, relative risk ratios can be estimated after controlling for the effects of pre-treatment differences between groups.

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.2 WERE PREDICTED SUBSTANCE ABUSE OUTCOMES OBSERVED? (CONTINUED)**

**Analytic Issues**

- There is a tendency for clients to have higher than normal substance use just prior to entering treatment. The pre-treatment period needs to be long enough to capture the typical consumption pattern of the client, i.e., longer than 2 or 3 months.
- Clients need to be compared against their own baseline data (usually the pre-treatment period), rather than combing the data across clients within a treatment condition and then comparing them to the aggregate base line data. Aggregating the data obscures how many used less, used more, or stayed the same. Consider using statistical tests that compare differences in scores for each client.
- Although abstinence post-treatment is the ideal, the evaluator is interested in measuring any change in the consumption of drugs (including alcohol) from pre-treatment to post-treatment. Measurement needs to be more sophisticated than use or no use of illegal substances. The consumption of specific illegal substances or classes of substances (including legal substances such as alcohol) needs to be measured. One way to measure consumption is frequency over a time period, e.g., number of times per month or per week. Another approach is number of days during a time period. It is important also to note if drugs are used in combination, e.g., heroin and cocaine as a speed ball.
- Since a client can substitute one drug or addiction for another post-treatment, it is import to ask about all drugs or classes during the periods of interest.
- Physical specimens are, for obvious reasons, generally preferred to self-reported substance use. However, physical specimens are more expensive and they cannot be used retrospectively beyond a limited time period. The evaluator can use a combination of self-reports and physical specimens. For those clients that have both self-reported data and a physical specimen for the same time period, the results from the physical specimens can be used to validate the self-reported data.
- Statistical techniques such as multiple regression (for outcomes that are continuous) or multiple logistic regression (for outcomes that are dichotomous) can be used to tease out the treatment effects from other influences. If more than one site is involved, then a hierarchical model can be used.

<b>EXHIBIT III-3 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>3.3 WERE PREDICTED SOCIAL AND FAMILY RELATIONSHIP OUTCOMES OBSERVED?</b>	
<b>Why Important</b>	<p>It is important for clients to establish healthy relationships with family members and social networks. Family problems may contribute to substance abuse, or they may be caused by substance abuse; consequently family issues are relevant to treatment. Substance abuse treatment services treat the whole client and not just his or her substance abuse.</p> <p>Substance abuse can be a multi-generational disease. Poor parenting skills are among the skill deficiencies of many substance abuse treatment clients. Many clients have child custody issues. Therefore, substance abuse treatment services may work on the parenting skills of their clients. Furthermore, adolescent clients usually live with their parents or surrogate parents, and there is a need to deal with the problems that lead to substance abuse or result from substance abuse.</p>
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did clients improve their relations with family members and significant others compared to their pre-treatment relations? Compared to control groups?</li> <li>■ Was there increased involvement of family members in treatment services compared to pre-treatment involvement? Compared to control/comparison groups?</li> <li>■ Did treatment service clients improve their parental functioning compared to pre-treatment functioning? Compared to control/comparison groups?</li> <li>■ Did treatment service participation decrease clients' involvement and/or risk of involvement with the foster care/child welfare system compared to pre-treatment? Compared to control/comparison groups?</li> <li>■ Did clients improve their housing and other safety and stability factors?</li> <li>■ Did children who participated in services with their parents demonstrate better social and family interactions?</li> <li>■ What client and SDU factors are associated with improved social and family relationships outcomes?</li> </ul>
<b>IEM tools</b>	<p><i>Staying in Touch: A Fieldwork Manual of Tracking Procedures for Locating Substance Abusers for Follow-up Studies</i></p> <p><i>Strategies for Follow-up Tracking of Juvenile, Homeless, and Criminal Justice System-Involved Substance Abusers: Overview and Bibliographies, 1990-1996</i></p> <p><i>Minimum Evaluation Data Set: Core Data Lists (section IV: Client Data)</i></p>

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.3 WERE PREDICTED SOCIAL AND FAMILY RELATIONSHIP OUTCOMES OBSERVED? (CONTINUED)</b>	
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Counselor interviews</li> <li>■ Client interviews</li> <li>■ Interviews with family members</li> <li>■ Evaluator observations of family interactions</li> <li>■ SDU records (including child protection services in certain circumstances)</li> <li>■ Family and juvenile courts records may be accessible in certain circumstances</li> </ul>
<b>Analytic Issues</b>	<ul style="list-style-type: none"> <li>■ Provide simple descriptions of post-only and pre/post comparisons of these issues. What services were provided to clients that related to family issues (e.g., help to obtain housing or family therapy)? Did all those who needed the service get the service? What were the impacts of these services on the relevant family variables? What was the relationship between participation and outcome? For example, did those who attended more family therapy sessions have better outcomes than those who attended fewer sessions? How did those who needed the service and did not get the service compare to those who did?</li> <li>■ Examine the outcome of child custody issues and its relationship to treatment outcome. Were those that had positive treatment outcomes more likely to have positive outcomes in their child custody cases?</li> <li>■ Although reuniting the family or improving its functioning are the ideal goals, they are not appropriate in all circumstances. Family members other than the client may have a drug abuse problem or be engaged in other problematic behavior such as criminal activity that makes reuniting the family inappropriate. The problems of family members other than the client may be outside of the scope of the services and improving the family functioning an unrealistic goal in these situations. In examining outcomes related to family behavior, the analyst needs to determine what is an appropriate outcome. This can be determined by examining the case record to ascertain the family circumstances and the goals set for the client by the counselor.</li> </ul>

<b>EXHIBIT III-3 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>3.4 WERE PREDICTED EFFECTS ON EMPLOYMENT AND EDUCATION OBSERVED?</b>	
<b>Why Important</b>	In addition to treating clients' substance abuse problems, treatment services need to help clients become productive citizens. Further, educational and employment problems can be a contributing factor to substance abuse. For most clients, gainful employment is a reasonable treatment goal. For those clients without a high school diploma or the equivalent of a high school equivalency diploma is a reasonable goal.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did treatment service clients increase their educational level (e.g., GED attainment) as compared to control/comparison groups?</li> <li>■ Did treatment service clients increase their employability (such as vocational skills, job search skills, or improved attitudes and behavior related to employment) as compared to their pre-treatment employability? As compared to control groups?</li> <li>■ Did treatment service clients increase their employment post-treatment compared to their pre-treatment employment? as compared to control/comparison groups?</li> <li>■ Did treatment service clients increase licit income compared to their pre-treatment income? compared to control/comparison groups?</li> <li>■ Did treatment service clients decrease reliance on welfare?</li> <li>■ Did children who participated in services with their parents demonstrate better educational achievements or outcomes?</li> <li>■ What client and treatment service factors are associated with improved employment and education outcomes?</li> </ul>
<b>IEM tools</b>	<i>Minimum Evaluation Data Set: Core Data Lists</i> (section IV: Client Data)
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Client interviews for self-reported data related to employment or education</li> <li>■ SDU records for information regarding employment or education</li> <li>■ School records</li> <li>■ Official records (e.g., welfare, unemployment insurance, etc.) for information related to employment.</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.4 WERE PREDICTED EFFECTS ON EMPLOYMENT AND EDUCATION OBSERVED?**  
**(CONTINUED)**

**Analytic Issues**

- Employment, educational services, and welfare are related concepts that need to be considered together in the analysis. Today with welfare reform, the goal is to have every able-bodied, mentally fit adult under retirement age gainfully employed. Although employment is the ultimate goal, there may be intermediate goals such as attending an educational or vocational services. Furthermore, attending an educational services usually precludes simultaneously being employed. It is important to examine change in the individual clients at least from pre-treatment to post-treatment. Individual characteristics such as health status, mental health status, educational attainment, and family structure affect employment and welfare outcomes.
- Employment is usually not a realistic goal for clients with severe medical or psychiatric disabilities or with little or no work experience or skills. It may be more appropriate to help the client apply for welfare or other benefits. Consequently, for these clients an increase in benefits would be considered a positive outcome.
- Since employment is often related to general economic conditions in the community, it is especially helpful to have a control, contrast, or a comparison group to account for the influence of this on post-treatment employment. It may also be possible to account for general economic conditions by different cohorts from the same SDU that had to find employment under different economic conditions.

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.5 WERE PREDICTED PHYSICAL HEALTH EFFECTS OBTAINED?</b>	
<b>Why Important</b>	<p>Many substance users have serious health problems; it is important that these problems be identified and dealt with appropriately. Many health problems involve physical pain for which drugs or alcohol are an escape and many health problems can prevent a client from attaining stable employment. Treatment service providers often seek to educate clients regarding these problems. Specific diseases of importance in this population are AIDS, hepatitis, tuberculosis, sexually transmitted diseases, and cirrhosis of the liver.</p>
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did clients comply with routine health maintenance visits?</li> <li>■ Did clients improve their overall health, taking into account pre-existing conditions (such as HIV infection, TB, STDs, and other infectious diseases)?</li> <li>■ Did clients experience decreased health problems that resulted in emergency room use, hospitalization, and doctor visits?</li> <li>■ Did clients improve their ability to engage in daily living activities?</li> <li>■ Did pregnant women who participated in treatment show a decreased rate of perinatal complications and improved birth outcomes (e.g., number of premature births, low birth weight, small size for gestational age, neonatal abstinence syndrome) as compared to control/comparison groups?</li> <li>■ Were children of pregnant women who participated in treatment born with fewer drug problems, as detected chemically when compared to control groups?</li> <li>■ Did children who participated with their parents demonstrate better health?</li> <li>■ Did clients decrease high-risk behaviors (for example, unsafe sex) compared to their pre-treatment behavior? as compared to control/comparison groups?</li> <li>■ What client and treatment factors are associated with improved health outcomes?</li> </ul>
<b>IEM tools</b>	<p><i>Minimum Evaluation Data Set: Core Data Lists</i> (section IV: Client Data)</p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Self-reported data regarding the health and health related behavior (including sexual behavior)</li> <li>■ Birth records (for information on births to women in treatment)</li> <li>■ Coroner reports (to determine cause of death and any secondary causes)</li> <li>■ SDU records (for information related to health, e.g., physical exam, referrals to health care providers etc.)</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.5 WERE PREDICTED PHYSICAL HEALTH EFFECTS OBTAINED? (CONTINUED)**

**Analytic Issues**

- Summarize client utilization of health care services received before and after treatment, particularly reflecting on whether there were increases in primary and preventive services received and decreases in hospitalizations, emergency, or urgent care visits. These are critical analyses for subsequent cost-benefit estimates.
- Issues related to health care access need to be explored. Clients may want to seek health care for their problems but may not be able to because affordable health care may not be available.
- Wherever feasible, multiple measures of health status should be taken. For example, in addition to diagnosed medical conditions, clients frequently have physical symptoms of other as-yet undiagnosed conditions.
- Clients' knowledge, attitudes, and behavior related to health care behavior (e.g., AIDS risk behaviors) needs to be measured at pre-treatment and compared to their post-treatment responses.
- Because clients often have better access to health care while they are in treatment, the health care status of clients post-treatment may appear to be worse than pre-treatment because undiagnosed problems were identified during treatment.
- The evaluator should bear in mind that data on health problems and health care may be culturally biased. Different cultures express more sensitivity to certain health topics (e.g., sexual practices), and language barriers may prevent clients from understanding medical terms.
- The impact of managed care needs to be considered in conducting these analyses. For example, hospital admissions may be influenced by managed care policies.
- It may be useful for the analyses to obtain information on the prevalence and incidences of various illnesses in the community.

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.6 WERE PREDICTED EFFECTS ON MENTAL HEALTH OBSERVED?</b>	
<b>Why Important</b>	Many substance users have serious mental health problems that, if left untreated, may maintain or exacerbate substance abuse and related problems. Hence, many substance abuse treatment services integrate mental health services for “dual diagnosis” patients.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did clients improve their overall mental health status via improved self-esteem and reduced anxiety, stress, depression, and suicidal tendencies as compared to their pre-treatment status? As compared to control groups?</li> <li>■ Did dually diagnosed clients show reduced symptomatology as compared to their pre-treatment symptoms? As compared to control/comparison groups?</li> <li>■ Did clients show increased self-esteem as indicated by standard instruments as compared to their pre-treatment ratings? As compared to control groups?</li> <li>■ What client and treatment factors are associated with improved mental health outcomes?</li> </ul>
<b>IEM tools</b>	<i>Minimum Evaluation Data Set: Core Data Lists</i> (section IV: Client Data)
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Self-reported data regarding mental health.</li> <li>■ Evaluator observations (interviewers can observe the respondent during the interview and note if he or she displays any psychiatric symptoms).</li> <li>■ SDU records (for information related to mental health, e.g., psychiatric examination, psycho-social histories, referrals for psychiatric treatment or indications of any psychotropic medications).</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.6 WERE PREDICTED EFFECTS ON MENTAL HEALTH OBSERVED? (CONTINUED)**

**Analytic Issues**

- Limited diagnostic information on psychiatric comorbidity is a common challenge in substance abuse treatment, and brief indicators of psychiatric symptoms may be too general to produce observable changes from before to after treatment. It may be necessary for the evaluator to focus on changes in measures of mental health that relate to client's ability to cope and function in society.
- Care must be taken not to confuse symptoms or behaviors that are the result of drug-taking with those that are the result of psychiatric problems. For example, an admission to a psychiatric facility could be the result of taking a drug like LSD or PCP, or it could be due to a psychiatric condition such as depression.
- As with health care, there should be analysis of mental health services used. When symptoms are present, services should be accessed—reductions in services should only be considered a positive outcome if underlying psychiatric symptomatology also diminishes.
- Where possible, distinctions need to be made between different types of psychiatric problems as well as distinctions in severity.
- Wherever feasible, multiple measures of mental health should be used. For example, clients need to be asked about their psychiatric symptoms as well as mental health-related behaviors such as medications and suicide gestures.
- The impact of managed care needs to be considered in conducting these analyses. For example, hospital admissions may be influenced by managed care policies.

<b>EXHIBIT III-3 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>3.7 WHAT WERE THE EFFECTS ON CRIMINAL JUSTICE INVOLVEMENT?</b>	
<b>Why Important</b>	In addition to treating clients' substance abuse problems, treatment services need to help clients become productive citizens. The criminal activity that is associated with substance abuse needs to be addressed by the treatment service. Furthermore, criminal involvement post-treatment may lead back to substance abuse.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Did clients reduce their rates of involvement with the criminal justice system as compared to their pre-treatment behavior? As compared to control groups?</li> <li>■ Did clients reduce their rate of arrest for committing violent acts as compared to their pre-treatment behavior? As compared to control groups?</li> <li>■ Were clients less likely to be arrested and formally charged for any offenses as compared to their pre-treatment behavior? As compared to control groups?</li> <li>■ Did clients spend less time in jail or on probation after treatment?</li> <li>■ What client and treatment service factors are associated with improved criminal justice outcomes?</li> </ul>
<b>IEM tools</b>	<i>Strategies for Follow-up Tracking of Juvenile, Homeless, and Criminal Justice System-Involved Substance Abusers: Overview and Bibliographies, 1990-1996.</i> <i>Minimum Evaluation Data Set: Core Data Lists (section IV: Client Data)</i>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Self-reported data regarding criminal acts, arrests, pending cases, convictions, time incarcerated, and sentences received.</li> <li>■ Criminal justice records relating to arrests, convictions, and sentencing.</li> <li>■ SDU records (for information related to criminal behavior, e.g., arrests, convictions, time incarcerated, and pending cases).</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.7 WHAT WERE THE EFFECTS ON CRIMINAL JUSTICE INVOLVEMENT? (CONTINUED)**

**Analytic Issues**

- Multiple measures of criminal activity and multiple data sources should be used. For example, clients need to be asked about their criminal activity, arrests, convictions, sentences, pending cases, and time incarcerated. Arrests, convictions and sentences reported by the client can be compared to official records. Because a client is not clear what constitutes an arrest and arrests may occur in a jurisdiction other than the one from which the data was obtained, consequently, it is possible for the client to truthfully report more arrests than the official record indicates.
- Compare numbers of crimes committed, numbers of arrests made, and amount of time spent in jail or its equivalent from before to after treatment. These comparisons are necessary also for subsequent cost-benefit estimates.
- In analyzing criminal activity it is important to distinguish between violent crimes and other crimes, such as drug-related crimes (e.g., possession of drugs or drug-related paraphernalia, the sale of drugs, or driving under the influence of alcohol or drugs).
- When working with arrest and conviction information, the analyst should distinguish between level of severity, e.g., between misdemeanors and felonies. (Level of severity distinctions can also be made for self-reported criminal activity.)
- When working with arrest data the analyst should be aware that these data reflect official policies and the behavior of criminal justice officials as well as the activity of the client. An arrest may be inflated in seriousness so that it can later be reduced by the court and the final conviction may be for a charge that is less than the actual offense committed by the client. Generally, arrest data are the main focus when examining criminal justice data, because the conviction record could be incomplete because of pending cases. Whenever possible, time spent incarcerated should be excluded when calculating arrests per time period. Although individuals can be arrested while they are incarcerated, this considered to be a relatively rare event.
- Pre-treatment, post-treatment, and during-treatment measurement periods should be approximately the same if comparisons are going to be made of criminal activity across time periods.

<b>EXHIBIT III-3 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>3.8 WHAT IS THE RELATIONSHIP BETWEEN THE COSTS AND THE OUTCOMES OF THE SUBSTANCE ABUSE TREATMENT SERVICE?</b>	
<b>Why Important</b>	Evaluators must combine economic information with other quantitative insights within processes and outcomes of treatment services. Cost effectiveness analysis allows for a comparison of relative outcomes to costs across two or more interventions, such as the number of reduced drug days per dollar invested in treatment group A versus group B. Cost benefit analysis expresses outcomes in dollar terms (e.g., an avoided hospital stay has a monetary value) allowing for a summation of economic benefits, comparison of benefits to cost (investments have positive returns if benefits exceed costs), and a comparison of benefit/cost ratios across two or more treatment groups.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ What are the monetary values of the treatment service outcomes?</li> <li>■ What is the relationship between the costs and the treatment outcomes/benefits of the knowledge-generating substance abuse treatment service?</li> <li>■ What is the benefit-cost ratio for the knowledge-generating substance abuse treatment service?</li> <li>■ What are the relative costs and benefits for the knowledge-generating substance abuse treatment service as compared to other substance abuse treatment services?</li> <li>■ How do cost offsets correlate with reduced drug use?</li> <li>■ For what percent of clients are net benefits positive?</li> <li>■ What are the <u>non</u>-monetary costs and/or benefits that should be considered?</li> <li>■ What is the relative cost offset of standard versus knowledge-generating service?</li> </ul>
<b>IEM Tools</b>	<p><i>Adding “Value” to CSAT Demonstrations</i></p> <p><i>Substance Abuse Treatment Cost Analysis and Allocation Template</i></p> <p><i>Minimum Evaluation Data Set: Core Data Lists (section IV: Client Data)</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Client records</li> <li>■ System records for criminal justice and health care services</li> </ul>

**EXHIBIT III-3 (CONTINUED)**  
**ANALYTIC APPROACH**

**3.8 WHAT IS THE RELATIONSHIP BETWEEN THE COSTS AND THE OUTCOMES OF THE SUBSTANCE ABUSE TREATMENT SERVICE? (CONTINUED)**

**Analytic Approaches**

- Append a cost per case to outcome data sets, either using an average per client cost for all clients in a treatment unit, or apply client-specific cost information.
- To compare cost-effectiveness for two groups, first tally outcomes across all members of one group. For example, determine the net percent of clients who reported improved family social functioning for each group, or sum the total number of reduced drug days or avoided arrests based on comparisons with pre-treatment data. Then, divide this aggregated outcome for each group by the total cost of treatment for that group. The resulting ratio can be compared across groups: the ratio with the highest value achieves the greatest number of outcomes per dollar spent—and hence is more cost-effective.
- To calculate cost-benefit, you need to first determine which of the variables that impose social costs have had a significant reduction. Analyze changes from pre- to post-treatment for the following variables (or others that impose social cost):
  - Number of arrests
  - Days incarcerated
  - Crimes committed
  - Days hospitalized
  - Emergency room visits
  - Dollars of welfare received.
- For variables with a significant change (increase or decrease), estimate the cost associated with each event of that variable. For example, the average cost of a hospital day or incarceration day is likely available from your local hospital or jail, respectively; national data are available from the American Hospital Association and the Department of Justice’s Bureau of Justice Statistics. The cost associated with crimes committed can be estimated using Bureau of Justice Statistics reports on the costs of crime. The cost associated with police arrests is the total budget of police and courts divided by the number of arrests or cases processed.
- Multiply each cost factor by the number of outcome events before and after treatment. When summed across all clients, you can generate an economic impact before treatment and an economic impact after treatment. Subtracting the post-treatment impact from pre-treatment impact provides a dollar-weighted “index of benefits,” or net savings. The net savings is the economic benefit of treatment that can be compared with treatment cost.
- Individual level economic benefits and benefit/cost ratios can be viewed as indexes of outcomes. These indexes can be treated as dependent variables in numerous follow-on analyses, e.g., in studies of the relationship between treatment length or duration and economic benefits, or subgroup analyses.
- Additional benefit-cost analyses can be performed for the pre- to during-treatment phase.

## **4. APPLYING GOOD JUDGMENT IN THE USE OF DATA**

A final step in analyzing treatment service and outcome data is considering how the evaluation process itself may have affected results and how the results themselves stack up against prior research. This type of information is used by the evaluator to prepare caveats to the analysis or to downplay those conclusions that may be the most suspect.

Evaluation-driven conclusions deserve strong headlines when they are based on valid, reliable and complete measurements and a solid evaluation design. Moreover, findings that are consistent with prior research gain the most immediate credibility, though when well designed and managed, innovative studies that challenge prior results or present new insights, often carry substantial impact. Alternatively, conclusions based on designs that do not allow the evaluator to adequately control for exogenous influences or measurement strategies that miss important factors or fail to obtain data from sufficient numbers of clients should be accompanied with strong caveats.

The following discussion on measurement reliability and validity, data quality, and evaluation design is intended to help evaluators gauge, “post hoc,” the quality of the design and implementation of the evaluation.

### **4.1 Assess Reliability and Validity**

Data—be they quantitative data, such as scales or independent measures, or qualitative insights—must possess two qualities: reliability and validity. A measure is considered reliable if, in a given situation, it produces the same results repeatedly. For example, a rubber tape measure that gives a different result every time it is used would be useless. Validity refers to the degree that an instrument measures what it purports to measure. In the case of physical measurements this is usually obvious. It is not so obvious, however, when measuring social behavior or psychological constructs.

Reliability takes precedent over validity. A scale or an instrument cannot be said to be valid if it is not reliable; it could, however, be reliable, but be measuring something different than what it purports to be measuring. For example, an instrument intended to measure or assess aspects of personality (which is considered relatively stable over time) may instead be measuring mood (which can vary throughout the day).

## **Reliability**

A scale, index, or instrument is considered reliable if, in a given situation, it produces the same results repeatedly; the same evaluator gets the same result from the same respondent at different times, or different evaluators get the same result from the same respondent at the same time. Reliability of quantitative data is reported using a reliability coefficient, which ranges from 0 to 1. A value of 1 would indicate perfect reliability.

Measures of social behavior and psychological constructs are typically less reliable than physical measures. For example, a doctor's scale is a more reliable measure of weight than an IQ test is of intelligence. Yet there is considerable variation among social measures as well—IQ tests are more reliable than most diagnostic measures of mental illness or clinical assessments of alcohol and drug use. The effect of unreliability is to weaken and obscure true differences among clients or between different SDUs. Generally, unreliable measures make treatment services appear less effective than they actually are.

Reliability is measured in a variety of ways, depending on the characteristics of the measure. For self-report measures, **test-retest** reliability assesses whether a respondent will provide a similar answer to a question if the question is asked more than once within a short time period. For clinical ratings, **interrater** reliability assesses agreement among multiple raters using the same information at the same time. **Internal consistency** reliability assesses how well multiple items reputed to represent a single domain (e.g., attitude toward drug use) provide consistent scores. A general rule of thumb for internal consistency reliability is that Cronbach alpha's (the most popular procedure for measuring internal consistency) of .70 and above are acceptable for most evaluation purposes. Other forms of reliability also exist, but these three probably cover most of the reliability assessments done on measures of substance abuse treatment outcomes. Methods for computing reliability statistics depend on additional factors, such as whether a measure is categorical or continuous. One advantage in using standard instruments is that reliability measures are known. Readers wishing to know more about reliability can consult standard psychometric texts.

## **Validity**

A measure is valid to the degree that it measures what it purports to measure. To establish validity, the results produced by the measure must be compared with the results of some other, clearly defined criterion. To illustrate validity, consider the following:

To validate a new measure of anxiety, a researcher might begin by asking those scored as anxious if they would rate themselves as anxious. Then the researcher might ask other observers who know the clients well (e.g., clergyman, doctors, or teachers). If clients' and observers' ratings were consistent with the new measure's results, the researcher would begin to gain confidence that it was indeed a valid measure.

But some anxious people attempt to conceal their anxiety. They will not rate themselves as anxious and may be adept enough at concealing their inner state that observers fail to perceive it. So the researcher really needs some other more objective criteria to validate this measure. Anxiety often produces physiological symptoms (heart palpitations, high blood pressure, breathlessness, perspiration) or cognitive disturbances (confusion, memory errors). If the new measure is a valid indicator of anxiety, its results should show a correlation with standard behavioral indices of anxiety.

As with reliability, there are several different types of validity:

- **Construct validity** assesses whether the measure truly measures the attribute (construct) it is said to measure, e.g., attitude toward drug use.
- **Content validity** is achieved if the items/questions are a representative sample of the concept or domain (universe) they purport to measure. With outcome measures, content validity is often established by relying on experts to judge whether the items truly represent the domain.
- **Concurrent validity** assesses the extent to which groups that are judged to be different based on a test criterion show differences on other measures related to the criterion (such as the anxiety example above).
- **Discriminant validity** demonstrates that the measure is more highly associated with other measures related to the criterion than with other measures unrelated to the criterion. Face validity simply means that the measure appears “on the face of it” to measure that which it purports to measure. All these forms (and others) have been used in the validation of substance abuse treatment outcomes. Like other aspects of evaluation, the determination of measurement validity has a political dimension, particularly in the absence of a clear validation criterion. Rossi and Freeman (1993) note that the perceived validity of an outcome measure often depends on whether the

appropriate stakeholders accept a measure as valid. As with reliability, readers wishing to know more about validity can consult standard psychometric texts.

Developing an accurate instrument to be used in outcome evaluation can be a difficult and expensive process. An alternative to developing your own scale or instrument is to use one developed by someone else. Many standardized instruments that are available and appropriate for outcome evaluation are in the public domain and are free to investigators. The psychometric properties and limitations of these instruments are known. Using standardized instruments has the added advantage of facilitating comparisons across different studies. One of the most important instruments for research and evaluation in the field of substance abuse is the Addiction Severity Index (ASI). The ASI is a structured, 45-minute, clinical research interview designed to assess problem severity in seven areas commonly found in persons who abuse alcohol and other drugs: medical condition, employment, alcohol use, other drug use, illegal activity, family relations, and psychiatric condition (see McLellan et al., 1992). Other standardized self-report measures are commonly used in mental health evaluation, e.g., the Beck Depression Inventory and the Rosenberg Self-Esteem Scale. For more information on standard instruments used in drug treatment research see the National Institute on Drug Abuse's publication on *Assessing Drug Abuse Among Adolescents and Adults: Standardized Instruments* (National Institute on Drug Abuse, 1994).

#### **4.2 Review Data Collection Procedures and Assess Data Quality**

Data collection procedures need to be carefully planned and pre-tested to minimize the introduction of errors into the data. Furthermore, procedures should not bias the outcome of the study, nor interfere with the operation of the intervention. Indicating to prospective evaluation subjects that they are somehow special may influence their behavior. In field study of work behavior, it was found that simply studying the participants produced an effect independent of the intervention. The subjects of the study felt special, and therefore, worked harder due to this rather than the intervention. This became known as the Hawthorne effect (Roethlisberger and Dickson, 1939). The contrary is also true; poor results have been attributed to subjects finding out they are in the contrast or control group, or because they thought they knew the expectations of the researcher. In addition to avoiding biasing the data, the intervention should disrupt the operations of the intervention as little as possible.

Determine how data collection procedures were tested before they were implemented, and how data collection staff, including staff who work for the SDU, were trained in data collection procedures. Describe the level of support from all stakeholders in the evaluation and in

particular, the support of staff. Determine also whether intrusive data collection methods were used. It is particularly important to describe how clients were informed about the evaluation, whether their participation was voluntary or not. The evaluator has a moral (and in many cases a legal obligation) to properly inform the clients about the evaluation if there is any possibility of the evaluation harming the client so that the client can make an informed decision regarding participating in the evaluation. Intrusive procedures, a lack of training, and insufficient efforts to secure client's cooperation can undermine the quality of the data.

Determine how data quality was monitored. Data should have been checked for quality and accuracy. This should be done at every stage of data collection. For example, a sample of the clients can be re-interviewed to check the accuracy of the original interviews. As the data are entered into the database, the entry program can make logic checks on the data for consistency and out-of-range entries. The data can be entered a second time to catch any data entry mistakes. Finally, after the data are entered, the analyst can examine summaries of the data to look for anomalies. If client identifiers are collected with the data, the data need to be stored in a manner that assures security.

### 4.3 Assess the Evaluation Design

Describe the fundamental evaluation structure (i.e., selection of groups for study or comparison, definition of data collection "points" or stages, and specification of the data to be collected at each stage of a data collection process) is fundamental to the credibility of the results. There are various practical concerns (e.g., financial limitations, time constraints, data availability, programmatic structure and ethical considerations) that can limit choice of evaluation designs.

- **Assess the degree to which observed changes in the target participants are caused by the treatment services, and cannot be plausibly accounted for in other ways.** Evaluations that possess this quality are said to have *internal validity*. These designs account for factors that may influence potential outcomes. Client characteristics and external influences such as changes in the economic environment are examples of factors that may influence outcomes.

For example, if one of the outcomes hypothesized for clients in a substance abuse treatment service is increased likelihood of employment, and there is a downturn in the economy resulting in increased unemployment, then the low rate of employment post-treatment may have more to do with the economy than with the treatment service. Similarly, if an SDU is treating substance users who are still in the early

phases of substance use (i.e., they are not addicted), are the positive treatment outcomes due to intervention or the type of client? Typically, to assure internal validity a control or a contrast group is used.

- Describe factors that enhance or limit generalizability of the conclusions. Can the study be generalized to and across other persons, settings, and times? Assess, for example: How representative are the clients, how typical is the treatment service, and were there any interactions between client characteristics and treatment outcomes, (i.e., did select subgroups have different outcomes)?

For example, if an approach to treating substance abuse was tested on teenage boys was found to be successful could we generalize the results to teenage girls? To adult women? If all the teenagers in the study came from a small town can we generalize the results to urban adolescents? If an evaluator is only concerned about the results of a specific treatment service or intervention, then external validity may not be an issue. Furthermore, internal validity takes precedence over external validity; if the study is not credible then it can not be generalized to other people or places.

- **Assess whether statistical tests were appropriate for the data and sample sizes.** An under-powered test would likely fail to detect a treatment difference when there is one. An over-powered test may waste resources. In considering power, it is important to consider the size of relevant subgroups and the total number of statistical tests as well as the size of the effect that you expect to find. Using an inappropriate statistical test may result in invalidating the evaluation or failing to find an effect because the test was not sensitive enough.

By the time that the evaluator has accumulated data, critical decisions have already been made and implemented regarding evaluation design. What remains is for the evaluator to describe how the design selected either avoids or is still subject to threats to internal and external validity. A more complete discussion of research design is included in Appendix B. For more information on selecting an evaluation design see the NEDTAC publication *Selecting an Evaluation Design*.

Remember that the ideal evaluation design is the randomized experiment. In a randomized experiment, threats to external validity from differences in client characteristics are accounted for by randomly assigning individuals to either a treatment or a control (non-treatment group) or to different treatment groups. The random assignments generally control for differences in client characteristics, since by chance both groups should be roughly equal on both known and unknown client characteristics. The comparability of the groups can be tested by using statistical procedures to compare the groups on known characteristics, such as age. If a particular client characteristic is of concern, then that variable can be included in the random

assignment procedure. For example, if gender is of interest, then men and women can be assigned randomly so that there are equal numbers of men and women in the groups. The effect of changes in the external environment on treatment outcomes is accounted for by having a control or alternative treatment group. Since the external environment is the same for all groups, any difference in outcome variables can be attributed to a treatment or intervention effect.

Unfortunately, random experiments are often not feasible in treatment settings. A good quasi-experimental design may be the most that can be accomplished under the circumstances. The evaluator needs to select the most appropriate and most rigorous design for the situation. The two quasi-experimental designs long considered to best approximate true experiments in terms of strength of causal inference are the regression-discontinuity design and the interrupted time series design.

Moreover, remember strong evaluation designs incorporate appropriate measurement periods and collect data on the treatment service context and cost structure for each treatment group being studied. For each treatment group, the evaluation design should incorporate measurements at multiple points in a client's treatment process: intake, during treatment, exit, and follow-up, e.g., 1 year after treatment had started. The evaluation design should also build in appropriate assessments of the service delivery units, counselor characteristics, and their cost structures in order to support the range of analyses recommended in Chapter III. Evaluations that rely on the Integrated Evaluation Methodologies package, including the Minimum Evaluation Data Set (MEDS) and the Substance Abuse Treatment Cost Analysis and Allocation Template (SATCAAT), are most likely to satisfy these requirements.

### **Design Drift**

Conducting an evaluation is different than conducting an experiment in the laboratory; consequently, the evaluation may not go as planned (Austin & Associates, 1982). The evaluator does not have full control over the evaluation and real world problems such as administrative difficulties can impact on the evaluation. The difference between the planned and the actual evaluation is referred to as design drift. Design drift can be a serious problem if it results in a study that is substantially flawed compared to the original design. In selecting an evaluation design, the evaluator has to take into consideration what can go wrong. In general, simpler designs are easier to execute than more complex designs.

There are many possible problems that one may anticipate in the evaluation of a substance abuse treatment service. It is incumbent upon the evaluator to consider the following:

- How did the evaluation design help explain findings?
- Were measurement strategies and data quality sufficient and appropriate to support findings?
- How were data analyzed and how could alternative data analysis approaches affect interpretation of results?
- How did findings compare with prior studies and literature?

Guidance to assist the evaluator when considering these questions follows.

<b>EXHIBIT III-4</b>	
<b>ANALYTIC APPROACH</b>	
<b>4.1 HOW DID THE EVALUATION DESIGN HELP EXPLAIN THE FINDINGS?</b>	
<b>Why Important</b>	<p>Evaluation results are strengthened when a rigorous evaluation design is well implemented. The evaluator needs, however, to discuss threats to the validity of the evaluation and to consider the impact on interpretation of results.</p> <p>Sometimes, the quality of comparison groups deteriorates during the course of an evaluation. For example, in a setting with randomized assignment, staff may realize that a particular patient would truly benefit from an experimental service and seek an exception to the random assignment.</p>
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ How were treatment and comparison groups selected and clients assigned to groups?</li> <li>■ How much baseline information existed regarding standard services?</li> <li>■ How clear cut were the differences between knowledge-generating and standard treatment conditions?</li> <li>■ What were the measurement points selected and were these appropriate?</li> </ul>
<b>IEM Tools</b>	<i>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluation</i>
<b>Information Sources</b>	Evaluation Plan
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Consider, based on client characteristics, records of client assignments to different treatment conditions, or the panoply of qualitative insights that discuss what factors may weaken comparisons of treatment effects. Were experimental or different services uniformly provided to clients in one condition and uniformly <i>not</i> provided to clients in the other condition?</li> <li>■ Was there contamination of groups? When, where, why, and to what extent?</li> <li>■ Did apparent differences between treatment groups remain stable over time or did they get stronger or weaker?</li> </ul>

<b>EXHIBIT III-4 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>4.2 WERE MEASUREMENT STRATEGIES AND DATA QUALITY SUFFICIENT AND APPROPRIATE TO SUPPORT FINDINGS?</b>	
<b>Why Important</b>	Evaluations of smaller projects often find that data quality is easily compromised, limiting generalizability of results. Moreover, evaluators of true “innovations” often find that standardized tools were not sensitive enough to the complex set of factors changing in clients or treatment systems. The evaluator needs to disclose how data quality and specific measures selected contribute to the presence or lack of findings.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Were information sources appropriate?</li> <li>■ How were data gathered regarding (a) clients, (b) client-specific treatment services, (c) staff, (d) SDU, (e) the environment?</li> <li>■ What data collection instruments were used, and were these appropriate? How were they selected or developed? What testing took place with these instruments? What problems occurred in their use?</li> <li>■ How were data reliability and validity assured? What threats to reliability and data influence interpretation of findings?</li> <li>■ How were the data coded, keyed, and maintained?</li> <li>■ What steps were built in to ensure data quality?</li> </ul>
<b>IEM Tools</b>	<p><i>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations</i></p> <p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Evaluation records</li> <li>■ Data collection forms</li> <li>■ Staff input</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ Describe how information was collected during the course of the evaluation, including steps to train data collectors, data management and quality control procedures, and steps taken to assure that data were complete, reliable, and valid.</li> <li>■ Determine whether response rates suggest bias. Were there patterns in participant response or item response rates, such as by subgroup of client or by other characteristics of clients or treatment services?</li> <li>■ Assess from staff and client perceptions whether the information that was provided during the course of the evaluation seemed accurate or subject to bias.</li> </ul>

<b>EXHIBIT III-4 (CONTINUED)</b>	
<b>ANALYTIC APPROACH</b>	
<b>4.3 HOW WERE DATA ANALYZED, AND HOW COULD ALTERNATIVE DATA ANALYSIS APPROACHES AFFECT INTERPRETATION OF RESULTS?</b>	
<b>Why Important</b>	The evaluator should be able to describe the statistical manipulations that were made to a raw data set such that another evaluator would be able to replicate the same analysis given the same raw information. Often times, analytic decisions (e.g., regarding trimming values in a data set) are made “on the fly” – decisions that can influence results. As noted in Chapter II, sensitivity analyses are necessary to judge the impact of analytic decisions.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ Were sample sizes of sufficient power to support needed analyses?</li> <li>■ Did statistical tests and analyses appropriately control for exogenous factors?</li> <li>■ Did qualitative information sources corroborate quantitative information sources?</li> <li>■ How were missing data handled?</li> <li>■ How were cost-offsets estimated? How were dollar values assigned to criminal justice, health, and/or productivity effects?</li> </ul>
<b>IEM Tools</b>	<p><i>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations</i></p> <p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Evaluation records</li> <li>■ Data collection forms</li> <li>■ Staff input</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ In many ways, this is a broad summary issue that may present an infinite set of applications. For example, controlling for exogenous factors depends on the time and resources available as well as the quality of data available to use in statistical models. The evaluator should describe what steps were taken, and what additional analytic steps are needed.</li> <li>■ Assessing the match between qualitative and quantitative insights at this point in the evaluation process requires that the evaluator step back and think, in the big picture, did the entire panoply of outcome and quantitative service data make sense in light of more subjective or impressionistic evidence? If not, why not? This does not require the evaluator to discard results that do not match but only to seek explanations.</li> </ul>

<b>EXHIBIT III-4 (CONTINUED)</b> <b>ANALYTIC APPROACH</b>	
<b>4.4 HOW DID FINDINGS COMPARE WITH PRIOR STUDIES AND LITERATURE?</b>	
<b>Why Important</b>	Finally, the evaluator is responsible for couching evaluation results in the context of prior literature. Results that corroborate other findings are most readily credible; results that contradict other findings may suggest more investigation are warranted. New results (where there is not a wealth of literature) almost always point to additional analyses that are needed.
<b>Analytic Questions</b>	<ul style="list-style-type: none"> <li>■ How did actual results compare against prior studies and existing literature?</li> <li>■ Did expected results actually occur? If not, were there factors in the delivery of the knowledge-generating service that account for that?</li> </ul>
<b>IEM Tools</b>	<p><i>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations</i></p> <p><i>A Guide to Process Evaluation for Substance Abuse Treatment Services</i></p>
<b>Information Sources</b>	<ul style="list-style-type: none"> <li>■ Evaluation records</li> <li>■ Data collection forms</li> <li>■ Research, evaluation, and substance abuse treatment literature</li> </ul>
<b>Analytic Approaches</b>	<ul style="list-style-type: none"> <li>■ The evaluator should describe methods and summarize the results from prior similar studies. Aim for the most recent studies and studies conducted in comparable settings. Then move to related settings or back in time as needed.</li> <li>■ The evaluator should seek to compare findings and explain differences they obtain with prior studies based on the methods or the underlying service differences.</li> </ul>

## REFERENCES

In addition to the NEDTAC series the reader is referred to the following works:

- Anglin, M.D., & Hser, Y.I. (1992). Outcome evaluation research for the fourth triennial report to Congress. University of California, Los Angeles: Drug Abuse Research Center.
- Austin, M.J., & Associates. (1982). Evaluating your agency's programs. Newbury Park, CA: Sage Publications, Inc.
- Chelimsky, E. (1978). Differing perspectives of evaluation. In C.C. Rentz, & R.R. Rentz (Eds.), Evaluating Federally Sponsored Programs: New Directions for Program Evaluation 2 (Summer), pp. 19-38. San Francisco: Jossey-Bass.
- Chen, H.T. (1990a). Theory-driven evaluations. Newbury Park, CA: Sage Publications, Inc. 167-189.
- Chen, H.T. (1990b). Issues in constructing program theory. New Directions in Program Evaluation, Issue 47. San Francisco: Jossey-Bass Publishers.
- Cook, T.D., & Campbell, D.T. (1979). Quasi-Experimentation: Design & Analysis Issues for Field Studies. Boston: Houghton Mifflin Company.
- Dennis, M. (1994). Ethical and practical randomized field experiments. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), Handbook of Practical Program Evaluation. San Francisco: Jossey-Bass.
- Gerstein, D., Johnson, R.A., Harwood, H.J., Fountain, D., Suter, N., & Malloy, K. (1994). Evaluating recovery services: The California Drug and Alcohol Treatment Assessment (CALDATA). Sacramento, CA: State of California, Department of Alcohol and Drug Programs.
- Gerstein, D., & Harwood, H. (1991). *Treating Drug Problems*. Washington, DC: National Academy Press.
- Glantz, S.A. (1997). Primer of bio-statistics. New York: McGraw-Hill.
- Guba, E.G., & Lincoln, Y.S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage Publications.
- Hubbard, R., Marsden, M., Rachal, J., Harwood, H., Cavanaugh, E., & Ginzburg, H. (1989). Drug abuse treatment: A national study of effectiveness. Chapel Hill, NC: University of North Carolina Press.
- Kachigan, S.K. (1991) Multivariate statistical analysis. New York: Radius Press.

- Leffieri, D. (1992). A primer of research strategies in alcoholism treatment assessment. In J. Allen, & F. Caldwell, F. (Eds.). NIAAA Treatment Handbook Series 3. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- 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(3), 199-214.
- Mohr, L.B., (1992) Impact analysis for program evaluation. Newbury Park: Sage.
- Moras, K. (1993) Substance abuse research: Outcome measurements conundrums. In L.S. Onken, J.D. Blaine, & J.J. Boren, (Eds.) Behavioral Treatments for Drug Abuse and Dependence (pp. 217-248). NIDA Research Monograph No. 137. Rockville, MD: National Institute on Drug Abuse. 217-248.
- National Institute on Drug Abuse (1994). Assessing drug abuse among adolescents and adults: Standardized instruments. [Clinical Report Series, NIH Publication 94-3757]. Rockville, MD: NIDA.
- Orwin, R., & Goldman, H. (1995) Evaluation of alcohol, drug abuse and mental health services. In B. Levin & Petrila (Eds.), Mental Health Services: A Public Health Perspective. New York: Oxford University Press.
- Pietrzak, J., Ramler, M., Renner, T., Ford, L., & Gilbert, N. (1990). Practical program evaluation: Examples from child abuse prevention. Newbury Park, CA: Sage Publications.
- Roethlisberger, F.J., & Dickson, W. (1939). Management and the worker. Cambridge, MA: Harvard University Press.
- Rossi, P.H., & Freeman, H.E. (1993). Evaluation: A systemic approach. Newbury Park, CA: Sage Publications.
- Scriven, M. (1993). Hard-won lessons in program evaluation. New Directions in Program Evaluation, Issue 58. San Francisco: Jossey-Bass Publishers.

**APPENDIX A:**  
**INTEGRATED EVALUATION METHODS PACKAGE:**  
**A GUIDE FOR SUBSTANCE ABUSE TREATMENT**  
**KNOWLEDGE-GENERATING ACTIVITIES—EXECUTIVE SUMMARY**

Since its inception, the Center for Substance Abuse Treatment (CSAT) has provided Federal leadership to improve substance abuse treatment accessibility, effectiveness, and efficiency. CSAT's mission and activities have evolved from directly supporting treatment services to supporting knowledge-generating activities. This evolution is evident in the current Substance Abuse and Mental Health Services Administration policy on evaluation as described in *Evaluation Policy*, SAMHSA, 1995.

The need for an integrated model of evaluation and planning at SAMHSA is presented in "Evaluation in the Substance Abuse and Mental Health Services Administration," *Evaluation and the Health Professions*, by Marsh, Jansen, Lewis, & Straw, 1996. CSAT also supports site-specific, cross-site, and national evaluations that have provided experience with a wide array of evaluation design and implementation methods. These experiences further supported the need for an integrated evaluation strategy and led to the development of a comprehensive set of evaluation products, including concept papers, technical assistance (TA) materials, and analytic tools. Collectively, these products are referred to as the Integrated Evaluation Methods (IEM) Package. The IEM Package organizes these products within an evaluation framework that is designed to support CSAT knowledge development and application goals. The evaluation framework itself was constructed on the basis of accumulated experiences among internationally known treatment service evaluation professionals. The IEM Package reflects and incorporates evaluation experiences gained over the past decade.

**Evaluation Framework and the Integrated Evaluation Methods Package**

National evaluation experiences have reinforced the fact that substance abuse treatment evaluation involves a standard set of tasks that generally occur in the following order:

- **Planning the evaluation/knowledge-generating activities**, which includes selecting the substance abuse treatment issue, identifying the theoretical foundation for the intervention, determining knowledge development program goals and implementation approach, and setting the evaluation goals and objectives that determine the overall parameters of the evaluation

- **Selecting the evaluation design**, which sets forth the overall strategy for establishing the process and outcome evaluation questions, measurement approach, and generalizability of findings
- **Developing the data requirements**, which flow from the evaluation questions and measures and include: SDU, clinician, cost, and client data
- **Developing data collection instruments**, which are based on the data requirements and are developed or selected from an integrated inventory of instrumentation
- **Collecting the data**, which includes developing data management processes and tools (including quality control procedures) and conducting the data collection activities
- **Analyzing the data**, which involves multiple levels of comparison and is governed by an analysis plan
- **Reporting the evaluation findings**, which includes evaluation knowledge dissemination and application within the field.

The evaluation process outlined above provided a framework for the development of products related to these evaluation concepts and methods. Taken together, those products comprise the IEM Package.

### **Integrated Evaluation Methods Products**

CSAT requested the development of a series of evaluation concept papers, TA materials, and tools to support and operationalize each phase in the evaluation of substance abuse treatment knowledge-generating activities. These items are included in the IEM Package. The concept papers are based on theoretical evaluation research constructs that have been adapted to substance abuse treatment services evaluation and knowledge-generating activities. The concept papers primarily support the evaluation planning phase and address such topics as the self-adjusting treatment evaluation model, cost analyses, and performance measurement. The TA materials and tools include specific evaluation methods that have direct applicability to substance abuse treatment knowledge-generating activities. The concept papers and TA materials that constitute the IEM Package are listed and briefly described in Exhibit I.

## EXHIBIT I

### EVALUATION FRAMEWORK AND INTEGRATED EVALUATION METHODS PACKAGE

EVALUATION FRAMEWORK	INTEGRATED EVALUATION METHODS PRODUCTS
<p><b>1. Planning the evaluation/ knowledge-generating activities</b></p>	<ul style="list-style-type: none"> <li>■ <b>Integrated Evaluation Methods: A Guide for Substance Abuse Treatment Knowledge Generating Activities:</b> Concept paper that describes the development of an evaluation framework, evaluation concepts, and TA materials to support the framework.</li> <li>■ <b>Self-Adjusting Treatment Evaluation Model:</b> Concept paper that describes an approach for integrating evaluation findings within treatment operations so as to adjust and improve service delivery.</li> <li>■ <b>Building Team Capability to Fully Implement and Utilize the Self-Adjusting Treatment Evaluation Model:</b> Concept paper to assist treatment providers in building capabilities to integrate the self-adjusting treatment model within day-to-day operations and service delivery.</li> <li>■ <b>Adding “Value” to CSAT Demonstrations:</b> The What, How and Why of Cost Analysis: Concept paper on the need for and types of cost analyses for CSAT demonstrations and knowledge-generating activities. (The Lewin Group)</li> <li>■ <b>Performance Measurement for Substance Abuse Treatment Services:</b> Concept paper about the increasing importance of provider performance measurement and analyses and an explanation of the case-mix adjustment methodology.</li> <li>■ <b>Client Levels of Functioning as a Component of Substance Abuse Treatment Services Evaluation:</b> Description of the rationale and methods for assessing client level of functioning and recommended core LOF data elements that could help to measure the effectiveness of treatment services received.</li> <li>■ <b>Substance Abuse Treatment Evaluation Policy Notebook:</b> These materials are aimed at facilitating understanding of the SAMHSA policy for evaluation and federal regulations on client confidentiality and assisting evaluators to meet CSAT evaluation requirements.</li> <li>■ <b>Substance Abuse Treatment Evaluation Resource Notebook:</b> The notebook contains evaluation bibliographies and listings of organizations, hot lines, on-line data bases, and contact information for obtaining assistance in evaluating treatment services.</li> </ul>
<p><b>2. Selecting the evaluation design</b></p>	<ul style="list-style-type: none"> <li>■ <b>A Guide to Process Evaluation for Substance Abuse Treatment Services:</b> TA tool presenting purposes of process evaluation and the application of process evaluation methods to single site and multi-site treatment services.</li> <li>■ <b>Using Logic Models in Substance Abuse Treatment Evaluations:</b> TA tool describing logic model purposes and techniques for designing and planning the evaluation of treatment services.</li> <li>■ <b>A Guide to Selecting an Outcome Evaluation Design for Substance Abuse Treatment Evaluations:</b> TA tool describing overall strategies for developing evaluation questions, measurement, controls, validity/reliability, sampling, design effects, and generalizability of findings. (Battelle)</li> </ul>

**EXHIBIT I (CONTINUED)**  
**EVALUATION FRAMEWORK AND INTEGRATED**  
**EVALUATION METHODS PACKAGE**

EVALUATION FRAMEWORK	INTEGRATED EVALUATION METHODS PACKAGE
<b>3. Developing data requirements</b>	<ul style="list-style-type: none"> <li>■ <b>Minimum Evaluation Data Set (MEDS): Core Data Lists:</b> TA tool for developing a uniform set of variables and response categories for the service delivery unit (SDU), clinician, cost, and client evaluation measures.</li> <li>■ <b>Substance Abuse Treatment Cost Allocation and Analysis Template (SATCAAT):</b> User manual to analyze treatment costs by unit of service for an SDU. (Capital Consulting Corporation)</li> </ul>
<b>4. Developing data collection instruments</b>	<ul style="list-style-type: none"> <li>■ <b>Substance Abuse Treatment Services Evaluation Data Collection Instruments:</b> Data collection instruments that fully incorporate the MEDS and that have been field tested for validity and reliability, as follows: Service Delivery Unit (SDU) Description; Clinician Background and Practice Survey; protocols to collect Adult, Adolescent and Child (in treatment with parent) Client Data at Intake, During Treatment, at Treatment Discharge and Post Treatment; Adult and Adolescent Record Extraction forms; and a section on protection of human subjects and informed consent.</li> </ul>
<b>5. Collecting the data</b>	<ul style="list-style-type: none"> <li>■ <b>Staying In Touch: A Fieldwork Manual of Tracking Procedures for Locating Substance Abusers for Follow-up Studies (UCLA):</b> User manual to establish and implement client follow-up data collection systems and procedures.</li> <li>■ <b>Strategies for Follow-up Tracking of Juvenile, Homeless, and Criminal Justice System-Involved Substance Abusers: Overview and Bibliographies, 1990-1998:</b> Description of tracking techniques used to increase response rates for follow-up interviews with homeless and juvenile/criminal justice involved substance abusers.</li> </ul>
<b>6. Analyzing the data</b>	<ul style="list-style-type: none"> <li>■ <b>A Guide to Substance Abuse Treatment Evaluation Data Analysis:</b> Recommended methods and procedures for analyzing process, SDU, clinician, cost, and client evaluation data.</li> </ul>
<b>7. Reporting the evaluation findings</b>	<ul style="list-style-type: none"> <li>■ <b>Substance Abuse Treatment Evaluation Product Outlines Notebook:</b> Compendium of outlines for evaluation products including evaluation plans, interim evaluation reports, final evaluation reports, replication studies, case studies, and ethnographies.</li> </ul>

## **CSAT Evaluation “Stakeholders”**

Evaluation “stakeholders” are individuals, groups, or organizations that have a significant interest in how well a program or activity functions. (See P.H. Rossi, H.E. Freeman, & M.W. Lipsey, *Evaluation: A Systematic Approach, 6th Edition*, 1999.) Within the context of the IEM Package, CSAT evaluation stakeholders include CSAT senior managers, CSAT project officers, and CSAT grantees and contractors including treatment service providers, coordinating centers, study sites, site-specific evaluators, and national evaluators.

### **Utility of the IEM Package for CSAT Evaluation Stakeholders**

While the conceptual and TA materials were developed from the perspective of the site-specific and multi-site evaluator, the concepts and TA tools have important utility for CSAT managers, project officers, and treatment service providers. The stakeholder’s position determines the perspective and utility of the IEM Package concepts and tools. For example, a CSAT senior manager can use the IEM Package to acquire a comprehensive evaluation context for planning and funding the knowledge-generating activities, the project officer can use the IEM Package to ensure that GFA/RFP applications are complete and include a full complement of design, execution, and product components, and the site-specific and multi-site evaluators can use the IEM Package to ensure that evaluation designs, data collection plans, data analyses, and product development have a consistent evaluation framework and compatible data across program areas. The suggested utility of the IEM Package for CSAT evaluation stakeholders is summarized in Exhibit II.

## EXHIBIT II

### UTILITY OF IEM PACKAGE FOR CSAT EVALUATION STAKEHOLDERS

STAKEHOLDERS	ROLES AND RESPONSIBILITIES	IEM PACKAGE UTILITY
<b>SENIOR MANAGERS</b>	<ul style="list-style-type: none"> <li>■ Policy development</li> <li>■ Issue identification for KD&amp;As</li> <li>■ Grant/contract funding decisions</li> <li>■ Overall program management</li> <li>■ Sustainability</li> <li>■ Dissemination</li> <li>■ Long-term strategic planning</li> <li>■ Program designs</li> <li>■ KA activities</li> </ul>	<ul style="list-style-type: none"> <li>■ Comprehensive evaluation framework</li> <li>■ Comprehensive evaluation components</li> <li>■ Roles and responsibilities for local/national evaluators as well as CSAT/grantee staffs</li> <li>■ Guidance for evaluation designs and products</li> <li>■ Standardized evaluation measures</li> <li>■ Logic models for program and evaluation design</li> </ul>
<b>PROJECT OFFICERS</b>	<ul style="list-style-type: none"> <li>■ GFA/SOW development</li> <li>■ Grant/contract application review</li> <li>■ Grant/contract monitoring</li> <li>■ Knowledge-generating products</li> <li>■ Identification and replication of promising practices</li> <li>■ Technical assistance assessment</li> </ul>	<ul style="list-style-type: none"> <li>■ Guidelines for high-quality evaluation designs (process and outcome)</li> <li>■ Logic models for program and evaluation designs</li> <li>■ List of evaluation measures with instrumentation</li> <li>■ Guidelines for evaluation products</li> </ul>
<b>GRANTEES: STUDY SITES</b>	<ul style="list-style-type: none"> <li>■ Grant applications</li> <li>■ Project development, implementation</li> <li>■ Local evaluation management</li> <li>■ Local evaluation coordination</li> <li>■ Knowledge-generating product development</li> </ul>	<ul style="list-style-type: none"> <li>■ Evaluation plan outline</li> <li>■ Process and outcomes evaluation designs</li> <li>■ SDU, clinician, cost, and client measures</li> <li>■ Roles and responsibilities for grantee provider/evaluator staff</li> <li>■ Guidelines for evaluation products</li> </ul>
<b>GRANTEES: MULTI-SITE EVALUATORS</b>	<ul style="list-style-type: none"> <li>■ Grant applications</li> <li>■ Comprehensive evaluation designs</li> <li>■ Evaluation implementation: <ul style="list-style-type: none"> <li>– Data collection</li> <li>– Data analysis</li> <li>– Reporting evaluation findings</li> </ul> </li> <li>■ Evaluation product development</li> </ul>	<ul style="list-style-type: none"> <li>■ Evaluation concepts</li> <li>■ Logic models</li> <li>■ Evaluation designs</li> <li>■ Evaluation data requirements</li> <li>■ Data collection instrumentation</li> <li>■ Data collection process and procedures</li> <li>■ Data analysis</li> <li>■ Product development</li> </ul>
<b>NATIONAL EVALUATORS/SERVICES RESEARCHERS</b>	<ul style="list-style-type: none"> <li>■ Contract applications</li> <li>■ Comprehensive evaluation designs</li> <li>■ Evaluation implementation: <ul style="list-style-type: none"> <li>– Data collection</li> <li>– Data analysis</li> <li>– Reporting evaluation findings</li> </ul> </li> <li>■ Evaluation product development</li> </ul>	<ul style="list-style-type: none"> <li>■ Evaluation concepts</li> <li>■ Logic models</li> <li>■ Evaluation designs</li> <li>■ Evaluation data requirements</li> <li>■ Data collection instrumentation</li> <li>■ Data collection process and procedures</li> <li>■ Data analysis</li> <li>■ Product development</li> </ul>

*IEM products and other evaluation materials may be obtained from:*  
<http://neds.calib.com>

## **APPENDIX B: MEASUREMENT CONCEPTS**

In the social sciences, measurement is a problem. For the most part, concrete entities are not being measured. Some concepts are difficult to define in a precise way and are, thus, difficult to measure. Generally, the more concrete the focus, the better the measurement. Usually, client characteristics and frequencies of specific behaviors are the easiest to measure. For example, age, sex, ethnicity, and marital status are typically accurately captured. Even these simple variables, however, can pose problems if the questions that identify them are not accurately worded. For example, age can mean date of birth, age on the date of admission to treatment or highest age obtained during a particular year.

### **1. FOUR TYPES OR LEVELS OF MEASUREMENT**

There are four levels of measurement: nominal, ordinal, interval, and ratio. The level of measure influences the choice of the statistical test to be used in the analysis. The simple measures contain less information and, therefore, the statistical techniques appropriate to this level of measurement have less information with which to work and are consequently less powerful, i.e., less likely to detect there is a difference when one exists.

- **Nominal measures.** The simplest form of measurement is nominal which is simple classification. Nominal measures consist of two or more non-ordered and mutually exclusive categories (they are also referred to as categorical measures). Examples are gender, ethnicity, and reasons for discharge. Counting the number of cases that fall into each category creates a frequency distribution, which in turn can be displayed in a frequency table or in one or more types of graphs.
- **Ordinal measures.** Ordinal measures are also grouped into mutually exclusive categories, but unlike nominal variables they are ordered or ranked in terms of size. Order scales are like measuring three children without a measuring tape. You can say John is taller than Mary who is taller than Bob is. A common type of ordinal measure is ratings of client satisfaction—typically the client rates his or her satisfaction with services on a 3- or 5-point scale. Rankings of services provided to clients in order of importance is another example of an ordinal scale. However, ordinal measures do not contain information about the distance between consecutive categories, or whether the difference between, say, scores of 3 (neutral, neither like, nor dislike) and 4 (like) is comparable to the difference between scores of 4 (like) and 5 (like very much). Like nominal measures, data from ordinal measures are typically displayed in a frequency table or in one or more types of graphs.
- **Interval measures.** With interval measures, the distance between values is uniform and meaningful throughout the scale, so the difference between scores of 3 and 4 is

equal to the difference between scores of 4 and 5, or 49 and 50, throughout the range of the scale. Interval measures therefore support arithmetic operations of addition and subtraction, which ordinal scales do not, but they do not support multiplication and division. This is because the absolute values of the scale are arbitrary, with no true zero point. A score of 20 is 10 points higher than a score of 10, but not twice as high. Temperature measured in Fahrenheit or Celsius is often given as an example of an interval measure because the choice for zero for either scale was arbitrary. One can not say that 50 degrees F is twice as hot as 25 degrees F.

- **Ratio measures.** Ratio measures are like interval measures with one important difference: interval measures have a true zero point. Height, weight, and age are all ratio scales; so are some commonly used treatment outcome questions like the number of days drugs were used during the prior 30 days. Interval measures support multiplication and division as well as addition and subtraction, and a score of 20 is twice as high as 10. The distinction between ratio and interval measurement is largely irrelevant when selecting statistical procedures to analyze evaluation data. The distinction between nominal and ordinal, and ordinal and interval, are far more important.

## 2. PARAMETRIC TESTS

Statistical tests that use interval and ratio data are called parametric tests. Those tests that allow the use of nominal or ordinal data are called non-parametric tests. The non-parametric tests also make fewer assumptions about the populations from which the data is theoretically derived. Parametric tests are more powerful than non-parametric tests, i.e.; they can detect a true difference between groups with a smaller sample. Parametric tests, consequently, are preferable to non-parametric tests. Although, parametric tests require interval or ratio data, nominal or ordinal in certain circumstances can be treated as if it were interval data. A detailed discussion of statistical theory or procedure is too complex to be treated here. The reader is referred to a general text on inferential statistics.

## 3. MEASUREMENT ERROR

Although all measurement is prone to error, this is particularly true with self-reported measures. Most measurements, furthermore, are intrusive; i.e.; they involve the active participation of the respondent. Typically respondents are required to answer questions either on a questionnaire or posed by an interviewer and occasionally they are required to perform a task or provide a physical specimen. Because cooperation is needed the respondent can refuse to

cooperate or interfere, either intentionally or unintentionally with the data collection in other ways. Self-reported information is especially prone to this. Types of error or bias typically associated with self-reported data (i.e., data collected through interviews or questionnaires) are:

- Acquiescence bias—trying to please the researcher (or clinician if clinical data is being used) by giving the response the subject believes that the interviewer wants.
- Social desirability bias—the respondent tries to appear more positive or socially acceptable.
- Recall bias—respondents forget.
- Misunderstanding the question
- Answering untruthfully (lying).

Through carefully wording questions and instructions for respondents, training interviewers and record abstractors, measurement error can be reduced. All items, i.e., individual questions need to be pre-tested so that problems can be identified and corrected before the data collection begins.

There are other approaches to data collection in addition to client self-reports. Observations and ratings made by clinicians or evaluators are not dependent on candid, competent responses from reluctant or impaired respondents including clinical assessments. Official records, e.g., arrests, hospital records or school records, can be a good source of data. Another approach used in substance abuse evaluation is the collection of physical specimens. Urine and hair samples are examples of this. All these approaches to gather data have biases associated with them. For a discussion on self-reported drug use compared to physical tests see: *Validity of Self-Reports of Drug Use at Treatment Admission and at Follow-up: Comparisons with Urinalysis and Hair Assays*. Whatever the approach used, field testing the procedures and properly training the observers, raters or data extractors will usually reduce errors in data collected.

Individual items on an interview instrument, check list or questionnaire can be combined into a scale that measures a specific psychological or social construct. Questions concerning drug use, for example, can be combined with related questions to develop an index of drug use severity. Clients can then be measure on the severity of their drug problem. Any scale is only as good as the individual items that comprise the scale.

**The perspective offered in this document is solely that of the author(s) and does not reflect the policies or views of the Federal government, or any of its Departments or Agencies.**