
**Practical Considerations in Designing and Implementing Evaluation
Research for Addiction Treatment Programs**

John Fitzgerald
Lewis & Clark College, Portland, Oregon

Thesis Committee
Gordon Lindbloom, PhD, Carol Doyle, PhD, Peter Barbur, MS

December, 1998

Practical Considerations in Designing and Implementing Evaluation Research for
Addiction Treatment Programs

Thesis written by

John P. Fitzgerald

B.S., University of Oregon, 1989
M.S., Lewis & Clark College, 1998

Approved by

_____	Gordon Lindbloom, P.h.D., Thesis Chair
_____	Carol Doyle, P.h.D., Committee Member
_____	Peter Barbur, M.S., Committee Member
_____	Gordon Lindbloom, P.h.D., Director, Counseling Psychology
_____	Jay Casbon, E.d.D., Dean, Graduate School

Abstract

In the changing managed care environment, outpatient addiction treatment programs are increasingly being required to implement outcome evaluation research to demonstrate clinical efficacy to third-party payers. Until recently, such research has been scarce or non-existent due primarily to limited resources and a lack of research expertise. However, many treatment programs are now recognizing the need for evaluation research, and learning how it can be used to improve delivered services. This article reviews manuals that can assist with evaluation research, and reviews additional literature within the framework of nine fundamental evaluation steps. To better understand the implementation of these steps, they are then applied to a case study involving a pilot outcome evaluation at an outpatient addictions clinic.

Introduction

One of the most critical challenges facing the behavioral health care system today is how to treat individuals with addictions. Many studies have concluded that substance use disorders represent the most frequently occurring mental health problem in the United States (Kiesler, Simpkins, & Morton, 1991; Miller & Brown, 1997; Regier et al., 1990). If behaviors such as gambling, sex and use of the internet can be addictive (Carnes, 1991; Schaffer, 1996), then the need for effective treatments is even greater. Numerous periodicals, professions and models have focused on addictions treatment, but the mechanisms underlying success in changing addictive behaviors are still not well understood (Miller, 1997).

Within the myriad of treatment options, there exists strong evidence that certain approaches and methods are more efficacious and cost-effective when compared to other treatment alternatives or placebo (Miller, 1997; Prochaska, Norcross & DiClemente, 1994). Although no single approach has been shown to be superior, managed care organizations (MCO's) are increasingly requiring treatment providers to substantiate their treatment methods and results (Dornelas, Correll, Lothstein, Wilbur, Goethe, 1996; Rainer, 1996). As providers have different styles and methods for delivery of treatment, there is a need for a recognized method of how to conduct evaluation research and report client outcomes (Johnson, 1995). Although there are a variety of helpful resources that outline various methodologies, to date no standardized evaluation process has been widely accepted for use with outpatient addiction clinics. The intent of this article is to provide a basic template upon which evaluation research

Technical depth that goes beyond the other manuals reviewed. For a first-time program evaluation, the five-part manual may be overwhelming. Some drawbacks or considerations include the manual following a one year time-line, lack of discussion involving special populations (i.e. children, multicultural), and provided measurement instruments that may not be practical or cost-effective to implement in a treatment setting.

Program Evaluation: A Do-It-Yourself Manual for Substance Abuse Programs. Published in 1996 by the Department of Psychology at the University of New Mexico, this manual was written with two purposes in mind. One is to explain how program evaluation can help to achieve program goals, the other is to show how to conduct an effective evaluation in easy-to-understand terms. Advantages include a straight-forward, nuts and bolts workbook that boils down the essential ingredients of how to conduct a program evaluation into nine easy-to-read chapters. The manual summarizes many complex topics in enough detail to be of value, but attempts to avoid technical language as often as possible. Its primary audience is those who have little or no research experience. It includes a number of helpful worksheets with examples, and has an appendix of assessment instruments which are selected for ease of use and low cost. Additionally, it includes a chapter on assessing special populations that addresses language, culture, and limitations of assessment instruments. Drawbacks of this manual include lack of a complete case presentation from start to finish, minimal discussion regarding research resource requirements, and no mention of ethical and legal issues.

Developing State Outcomes Monitoring Systems for Alcohol and Other Drug Abuse Treatment (Treatment Improvement Protocol – TIP #14-series). Printed in 1995 by the U.S. Department of Health and Human Services, the manual's purpose is to assist single state agencies, direct treatment providers, and other interested research and health professionals in the development, implementation and management of an Outcome Monitoring System. Although step-by-step instructions are not as clear, it offers chapters on topics not covered in the previously reviewed manuals that include: political considerations, ethical and legal issues, and budget considerations. Since the results of evaluation research are often provided to state and federal funding agencies, MCO's, and other government organizations, it can be helpful to know up front what information they require, and their recommendations for conducting on-going outcome research. This manual addresses these issues, and offers helpful follow-up contacts that can provide valuable assistance. It is not written specifically for treatment providers, and is less

examples of data that should be collected and reviewed routinely. The primary goal of this step is

client populations, treatment revenues and costs, attrition, and resource allocation are all rather than a time-limited research study. Information including demographic characteristics of and Eisen & Dickey (1996) suggest that evaluation research be viewed as an ongoing routine and accomplishing what it set out to do (National Institute of Health, 1993). Dornelas et al. (1996) and available resources, since evaluation is ultimately used to assess whether the program is

treatment program. It is helpful to begin by identifying the program's mission, goals, objectives, A starting point with any evaluation is to *clarify what is already known about the*

increasingly requiring outcome data, they are often very willing to provide assistance.

a first-time evaluation such as those already reviewed. Additionally, because MCO's are number of guidebooks, manuals and software packages that are invaluable when conducting this article and others will help to explain basic terminology, definitions, and steps. There are a

The first step is to *develop a general understanding about evaluation research*. Reading

process.

each step is provided here, focusing on practical considerations in the design and implementation "making sense of your data," "analyzing your data," and "data analysis." A brief discussion of

variants in wording. For example, step eight involved analyzing data, and labeling included:

different models all had nine steps, with each step in most cases being the same except for

create a template of the fundamental evaluation steps (Figure 2). Of the models evaluated, three that the steps involved in evaluation research are basically the same. Therefore, it is possible to

which one is most appropriate. Fortunately, after reviewing the most popular models, it is clear

literature yields a variety of methods and models, putting the burden on the evaluator to choose

could follow when conducting evaluation research. Unfortunately, a quick review of the

It would be ideal to have a standardized methodological process by which all clinicians

Review of Literature within the Framework of the Fundamental Evaluation Steps

review other resources as well.

provide invaluable assistance for the neophyte researcher, but it can also be extremely helpful to

The above manuals can all be obtained for free or a minimal charge. Used together, they

instruments, or discussion on data analysis techniques included in the manual.

suitable for a first-time project. There are limited worksheets, no case studies, assessment

to develop a core data set that describes the clientele served (i.e. gender, race, age, etc.), allows for analysis of "what happens" to clients during treatment (group vs. individual), and evaluates the outcomes of the population, or subgroups within the entire population (successful completion vs. dropout). Gathering this information can also become part of a continuous quality improvement program, where evaluation results become a benchmark for "improvement" rather than a management tool used to assure productivity (Johnson & Shaha, 1996).

The third step is to *select the type of evaluation*. There is a wide range of terminology and definitions used to describe evaluation research, and various terms often describe the same type of evaluation causing confusion when gathering information from various resources. In addition, terms such as "outcome" often have been used in so many ways that the definition becomes too broad for clinical usefulness (Schlosser, 1995). To help sort out the confusion, Figure 1 provides a schematic representation of the various types of evaluation research. A definition of evaluation research is a "methodical procedure for determining the relative impact of individual program components, component combinations, or the whole program, on the targeted individuals or groups" (Miller & Wolfe, 1996, p. 11).

The two primary types of evaluations are *summative* and *formative*. Summative evaluations focus on the big picture. They answer questions including: How well is the program working? What impact is treatment having on clients once they leave the program (i.e. outcome)? How does the program compare to other programs (i.e. efficiency)? In financial terms, summative evaluation looks at the bottom line (i.e. cost-benefit). Is the program profitable and worth continuing? Is the program meeting its mission and objectives in the most efficient manner? Alternatively, formative evaluations focus on the process of how a program can be improved and made more cost-effective. Instead of examining the big picture or the forest, formative evaluations investigate the parts and look at the trees. They answer questions including: How can retention of clients be improved? What portion of the target population are enrolling for treatment (i.e. reach)? How can expenses be reduced while maintaining or improving the level of service provided (i.e. process)?

Both types of evaluations are equally important, and often can be addressed in the same research project. Whether conducting summative or formative evaluations, Figure 1 provides a further breakdown of subtypes of evaluations. It is within the subtypes that multiple terms are used in the literature to describe identical types of research. It should be noted that the term

selected to describe a particular type of evaluation is not as important as understanding what that term means and how it is being used.

The next step is to *select the evaluation question(s)*. The question(s) desired to be answered follow from the type of evaluation chosen. As discussed previously, in general, summative questions focus on the results of a program, where formative address the process. Regardless of which questions are asked, this step involves restating the general questions into terms that can be measured, or operationalized. For example, a summative question may ask "does outpatient addiction treatment have an impact on sobriety time?" Restating it in measurable terms the question may be "what percentage of clients who have completed outpatient treatment have maintained their sobriety after six months?"

The fifth step is to *select the evaluation design*, which is the overall structure or recipe of the study. The National Institute of Health (1993) suggests that a design ideally should provide evidence that clients changed in some important way between intake and discharge (and/or follow-up), and that the change can be attributed to the treatment program. There are four basic evaluation research designs: posttest-only, pretest-posttest, experimental, and static group comparison. Posttest-only designs simply measure some variable (i.e. client satisfaction with the program) following treatment. This design has no baseline (or pretest), so results cannot be used to assess change over time. Nor does it allow a researcher to say that the results were caused by treatment. It is the simplest design, and can provide valuable feedback depending on what is being evaluated. The next design, pretest-posttest, has the benefit of measuring change over time. It compares values at intake with values at discharge or other points in time. This design has the benefit of determining whether change took place, but cannot be used to conclude that observed changes were caused by treatment. To show causation, an experimental design should be used. This is the optimal design and involves randomly assigning participants to a treatment group and a control group. It is optimal because the design controls for most threats to validity (variables not controlled for including chance events, maturation, and history). If changes are noted in the treatment group and not the control group, there is strong evidence that treatment was the cause for the change. However, in addition treatment programs it can be extremely difficult to implement since it would be unethical to place some clients in a control group that receives no treatment, or less than effective treatment. The final design type, quasi-experimental/non-equivalent control group, is used when a clinic wants to determine if some new treatment is

effective. For example, a clinic may have eight groups, and decide that four will receive a new treatment component, while the other four groups continue to get standard treatment. The results are then analyzed to determine if the treatment had any significant impact. The disadvantage is that participants for the study are not randomly assigned and noted changes in the groups may be due to participant attributes instead of treatment. However, pretest measures can be done to show group similarity which can increase the validity of the results. Usually, the design of an evaluation will be based on the evaluation question(s). If showing change over time is the desired goal, then either a pretest-posttest or quasi-experimental design will be sufficient. If it is important to show that change is the result of treatment, then the ideal design will be experimental. In this case, it is important to note again that due to the "real-world" nature of research, it is not always possible to utilize an experimental design when conducting evaluation research.

Once an evaluation design is chosen, it is necessary to *select the measurement tools* that ideally are standardized. This means research has been done to establish the measure's reliability, validity, sensitivity, interpretability, and norms for various populations (Graham, 1994). Since there is an unlimited number of assessment instruments available, it is important to select measures carefully. Pheiffer, Soldivera, & Norton (1992) have suggested four general principles for choosing instruments: 1) practicality, or ease of use, 2) sensitivity to change, 3) suitability for the client population, and 4) psychometric quality. Graham (1994) adds to these considerations the need for measures to be: valid (ability of a test to measure what it was designed to measure); acceptable (ethical – i.e. does not involve deception or manipulation); clinically useful; brief and easy to administer; able to provide an overall score that is a meaningful measure of status, and ; available for use at no cost or minimal cost. Measurement tools may include client questionnaires, structured interview data, standardized assessment instruments, biochemical markers of substance abuse (such as urine tests), and clinical observation (Miller & Wolfe, 1996). The easiest measures to use in a study are published, standardized assessment questionnaires that have been created to meet many of the previously discussed guidelines. There a number of manuals and handbooks that have summaries of widely used tools and their applicability to various types of research designs.

The seventh step is to *collect the data*. The core data set, along with measurement tools normally dictate the type of information being collected. Most evaluation manuals suggest the

use of a personal computer to track and record data. Simple spreadsheets or databases can be used to code, format and manage data effectively. Since ethical guidelines strictly protect the confidentiality of clients, it is necessary to obtain a signed, informed consent from each participant prior to commencing the study (Dornelas et al, 1996). Additionally, all collected data should be stored safely (computers should be password protected) and preferably coded using numbers instead of names (U.S Department of Health and Human Services, 1987). The three most important points at which to collect data are (1) when a client enters the program (intake), (2) completes the required treatment (discharge), and (3) at some designated period(s) after discharge. Since research on substance abuse treatment outcomes indicate that clients relapse within the first six months following program completion, the appropriate time to follow-up with clients is at 3, 6 and/or 12 months intervals (McLellan & Durell, 1996; Miller & Wolfe, 1996). The most common follow-up methods include phone interviews, mailed questionnaires and face to face contact. Collecting follow-up data can be extremely challenging when clients move, change phone numbers or have relapsed and are embarrassed to provide follow-up data. Suggestions for improving follow-up rates include: 1) presenting the program at intake as involving both treatment and follow-up, 2) having clients pay a monetary deposit that is refunded after collection of follow-up data, and 3) having clients provide two or three additional contacts that could be utilized to find clients should their phone or address change.

The eighth step is to *analyze the data*. This should be a rewarding experience because it involves answering the evaluation question(s). Since data analysis can be intimidating for those with limited statistical background, most articles and manuals suggest enlisting the assistance of a statistician or graduate student if necessary. However, like most other steps, analyzing the data can be simple or very complex depending on the type and need of analysis. A fair amount of analysis can be done simply by following steps in a guidebook. The process involves describing the data using descriptive statistics, and analyzing the results using inferential statistics. Describing the data normally includes calculating the mean, median and mode on variables including the demographic variables such as gender, age, time in treatment, and outcome. Variability should also be calculated using the standard deviation. These measures alone are often sufficient for answering many evaluations questions (i.e. descriptions of clinic populations). However, inferential statistics provide the added benefit of determining whether the difference observed between groups or data sets is statistically significant; meaning that the

observed outcome has only a slight chance of randomly occurring. Inferential statistics allow a researcher to make the assumption that the results from the sample population (i.e. those who participated in the study) can be generalized to the entire treatment population. The most common significant tests include: the t-test, analysis of variance, and Pearson correlation. The final step is to *report the findings*. Generally, results are provided to clinic employees, administrators and managers, and other stakeholders who would benefit from the evaluation. It is among the easiest steps because it usually follows a predetermined formula. The parts of the report include: abstract or executive summary, introduction, methods, results, and discussion. Miller & Wolfe (1996) suggest an abstract be a paragraph long and answer four basic questions: 1) what questions did you address? 2) what did you do to answer the questions? 3) what did you find? and 4) what do you conclude about your findings. The purpose of the introduction is to describe why the research is being conducted and what questions it hopes to answer. It often includes predictions for outcomes and an overview of the topic studied. The methods section addresses how the research was carried out step-by-step. It discusses the participants selected for the study, the measurement tools used, and the procedures that were carried out to complete the study. The results simply describe what was found. There is no attempt to interpret or explain the results. This is done in the final discussion section of the report, sometimes called the conclusion. It discusses the meaning of the results, steps that should be taken based on the outcome, weaknesses of the study, and suggestions for further research. This final step is very important to addiction treatment programs, since evaluation reports on core data sets and outcome results are extremely persuasive when applying for state or federal contracts, grants, or accreditation programs.

To better understand these steps, they will now be applied to a pilot outcome study conducted in an outpatient addiction clinic. Like the majority of clinics providing addictions treatment, the clinic used for the pilot study had never conducted any type of evaluation research. A first-time evaluation is ideally introduced through the use of a pilot study (Dornelas et al., 1996), which provides an opportunity to work through the previously discussed steps and gain knowledge of the evaluation process. It also can help identify problems that may arise before evaluating an entire program or population. In the next session, a brief overview of the clinic is provided (step two), followed by the specific methods, results and conclusions of the pilot study.

Pilot Study

The Oregon Health Sciences University Behavioral Health Clinic (BHC) has a 50-year history of delivering outpatient addiction treatment services to a broad base of clients throughout Oregon. It was formally the Addictions Treatment and Training Clinic, but in 1998 became the BHC to reflect the increased scope of mental health and dual-diagnosis services. Located conveniently in downtown Portland, BHC is a stand-alone clinic that employs an average of eighteen staff. The clinic only employs professionally trained clinicians who have as a minimum educational requirement a masters degree in a counseling-related field. The clinic mission is to create positive change in behavioral health by offering a supportive, therapeutic environment, where clients have access to a multitude of treatment approaches and styles.

Treatment services are provided to three populations: Clinic, Driving Under the Influence Intoxicated (DUI), and Gambling. Clinic clients may be self-referred, or come to the clinic at the request of a physician or friend. They typically have substance use disorders that may or may not be complicated by other mental health diagnoses. This population includes both mandated (criminal justice system) and non-mandated clients, homeless individuals, terminally ill clients, and those who may be referred for pain management when medications become abusive. DUI clients are mandated to enter the clinic for diversion treatment (usually first offense) or conviction treatment (multiple offenses). This population may or may not meet the criteria for a substance use disorder, and may enter the Clinic group after DUI treatment should they need further assistance with issues outside of the DUI treatment requirements. The gambling program is funded from the state gambling commission, and is for clients with pathological gambling problems, which may or may not have comorbid mental health disorders. Due to difficulties in obtaining a sufficient sample size, only clinic and DUI clients were enrolled in the pilot study.

All clients entering the program begin by having an orientation, intake evaluation and treatment planning session which includes the formulation of a treatment plan. During this process, the client's treatment strategy is mutually agreed upon. The program's treatment philosophy is based on the biopsychosocial model. Biological aspects of addiction that are addressed by treatment include genetic predisposition, pharmacological interventions, and psycho-medical issues that complicate treatment such as chronic pain and head injuries. Psychological aspects include interventions focused on thoughts, emotions and behaviors. Additionally, attention is given to the discussion of religious and spiritual issues as they relate to

the treatment process. Social aspects include a client's living environment, relationships and 12-step involvement. Although 12-step programs are seen as a positive adjunct to treatment, the clinic's program is not based on a 12-step model. Modes of treatment include individual, group, family and couples counseling.

Methods

The methods utilized for the outcome pilot study followed the fundamental evaluation steps previously discussed. An investigation was done to determine what data was currently being collected by the clinic. The result was the creation of a core data set developed to track information regarding: type of clients, revenues, gender, race and other variables (Table 1). This provided an overview of the clinic clientele, but to gain another perspective intake testing results were also tracked and became part of the core data set (Table 2). The testing instruments used during intake were the Beck Depression Inventory (BDI), the Substance Abuse Subtle Screening Inventory (SASSI), and the Folstein Mini-Mental Status Exam.

The Beck Depression Inventory is among the most common instruments for assessing depression. It is a 21-question measure that takes about 10-15 minutes to complete. A score of 0-10 indicates no depression to mild depression, 11-20 indicates moderate depression, and a score of 21 or above is evidence of significant depression. The SASSI is a 15-20 minute test that evaluates both alcohol and drug history from first use to present. It classifies each person as having a low or high probability of having a substance dependence disorder. The Folstein Mini-Mental Status Exam is one of the most widely used mental status tools used to assess overall cognitive functioning. It takes about 10 minutes to administer and contains 30 scorable items having to do with memory, attention and language production. A score below 27 indicates the need for further assessment.

Each of these instruments are well-known measures with high reliability and validity. They were considered as potential measures for the outcome study, but they also involved more time in administering, collecting and scoring that was beyond the scope of the pilot study. Additionally, in order to measure change between intake, discharge and follow-up, it would have been necessary to administer the tests at discharge which at the time of the study was not being done. Since one purpose of the study was to utilize data collected under current clinical practices, the test results are included to provide another view of the clinic demographics.

- Since the purpose of the pilot study was to begin an on-going evaluation to investigate the outcome of treatment for clients one month after discharge, the study was a summative, following the evaluation steps, the next task was selecting the evaluation question(s) which proved to be challenging. The only information being collected both at intake and discharge was data on the American Society of Addiction and Medicine's (ASAM) criteria for alcohol and drug placement. A copy of the intake packet that details information collected under each dimension can be found in Appendix 1. The clinic follows the ASAM criteria because they are mandated by the Oregon Administrative Rules. They assess a client on six separate dimensions: (1) detox/withdrawal (2) physical health conditions and complications (3) emotional/behavioral issues (4) treatment acceptance/resistance (5) relapse potential (6) recovery environment. From these dimensions clients are placed in one of four levels of treatment: 1) outpatient, 2) intensive outpatient, 3) residential, or 4) medically managed (hospital). BHC only provides level 1 and 2 service.
- At the time the pilot study was being designed, it seemed appropriate to develop one question that would assess changes in each dimension that could be compared to data collected at intake and discharge. Questions were developed by the research staff after reviewing the information collected at intake using the intake packet, and ASAM summaries at discharge. The goal in designing each question was to capture the most salient information under each dimension. The original questions proposed were:
1. Dimension 1: Have you had any alcohol or drugs in the last 30 days? If yes, what *quantity* and *frequency*?
 2. Dimension 2: Have you experienced any changes in your *physical* health in the last 30 days?
 3. Dimension 3: Have you experienced any changes in your *emotional* health in the last 30 days? (This includes depression, anxiety, etc.)
 4. Dimension 4: What *stressors or pressures*, including any *arrests*, have you experienced in the last 30 days?
 5. Dimension 5: What, if anything, did you learn in treatment that has changed your use of alcohol or drugs in the last 30 days?
 6. Dimension 6: Have you had any changes in your *recovery environment* in the last 30 days? (This includes friends, support groups, housing, etc.)

To enable comparisons across time, each question needed to be operationalized. Question one was easily measured by days of use in the last 30 days. However, questions two through six were more ambiguous and needed clarification before data collection. For question two, changes in physical health were defined as health issues needing current treatment. Qualifying problems included: skin rash, high blood pressure, bleeding ulcer, hepatitis C, asthma, migraine headaches, diabetes, and sexually transmitted diseases. Ideally, collecting data on specific problems would have allowed for more detailed analysis, but for the pilot study answers were coded on a yes/no

(This includes friends, support groups, housing, etc.)

6. Dimension 6: Have you had any changes in your *recovery environment* in the last 30 days?
5. Dimension 5: What have you done to help maintain your abstinence during the last 30 days?
4. Dimension 4: How accepting have you been of your current life issues and problems in the last 30 days?
3. Dimension 3: Have you experienced any changes in your *emotional* health in the last 30 days? (This includes depression, anxiety, etc.)
2. Dimension 2: Have you experienced any changes in your *physical* health in the last 30 days?
1. Dimension 1: Have you had any alcohol or drugs in the last 30 days? If yes, how many days out of the last 30 did you drink or use?

Unfortunately, in the preliminary stage of data collection, it became obvious that attempting to condense each ASAM dimension down to one question was problematic. Assessing each dimension at intake involved asking multiple questions, so asking only one question appeared to not address each dimension adequately. However, the purpose of the pilot study was to get a sampling of results for each dimension, and then conduct follow-up research in more depth where the results indicated a need. To allow comparisons across time, and obtain more specific information, the questions were revised. The revisions included limiting the data collected on dimension one by eliminating quantity, and only asking about days used since discharge. Additionally, question four was revised to reflect more accurately that information that was collected at intake and discharge. Question five was also revised since in its original form it could not be compared over time. The revised questions became:

basis for simplicity. For question three, emotional and mental health issues were defined by a positive DSM IV diagnosis, excluding substance use disorders. At follow-up the goal was to determine whether the patient still had mental health problems, and answers were coded again on a yes/no basis. Due to limited time and resources, efforts necessary to fully diagnose mental health disorders at follow-up was not done, limiting the specificity and objectivity of the data collected. Question four was also collected on a yes/no basis and criteria for acceptance of life issues or problems were based on the counselor's interpretation of the client's feelings or reactions at intake, discharge and follow-up. Responses to the question were grouped into two categories; those who accepted their life issues and problems and need for treatment, and those who did not believe they had a problem or needed treatment (primarily mandated clients). The pilot study did not compare mandated/non-mandated clients with those who were accepting or not accepting of life issues/problems, but this should be done in a complete evaluation. Question five was interested in client's risk for relapse. Since most counselors assessed risk subjectively on a low/moderate/high scale at intake and discharge, this scale system was also used at follow-up. Determination of risk was based on the client's current level of recovery activity, or in the case of DUII clients, whether steps were still being taken to avoid further legal problems from drinking and driving. For question six, recovery environment included a client's housing, support groups, friends and family, and other identified environmental factors that influenced recovery. Information was collected on a yes/no basis, and whether the changes were positive or negative were not tracked in the pilot study. Clearly, follow-up data collected for questions two through six involved a great deal of subjective interpretation by the interviewer.

Since results from the above questions were compared to intake and discharge data, the study followed a pre-post design. As discussed previously, the intake assessment instruments were considered as measures for the study. However, since they were not currently being used at discharge, and it was beyond the scope of this study to readminister them, the responses to the created ASAM questions became the measure of outcome. Using the original intake packet at discharge and follow-up would have added more validity and reliability to the results, but this was not possible due to the time involved in collecting all of the information included in the intake.

A sample of twenty-nine clients were selected to participate in the pilot study. The subjects were recruited from the entire population of clients during a five month period in 1998

(July – November) and asked to participate in the study either at intake, during treatment, or at discharge. All clients were told that the purpose of the study was to help the clinic learn what impact treatment had at least thirty days after completion of the program. They were provided a written consent form and all agreed to have their progress monitored after discharge. All clients who entered treatment at the clinic completed a two-hour intake evaluation with a professional counselor that included testing and an intake evaluation. The results of the evaluation were reviewed with each client and a mutually agreed upon treatment plan was developed, followed by treatment. At discharge, each client was again assessed by a counselor using the ASAM criteria. The assessment process involved reviewing the client's progress while in treatment, and subjectively summarizing progress made on each dimension in a written report. Discharge dates were tracked and after 30 days attempts were made to contact clients either by mail or phone.

When contacted by the researcher or a supporting staff member, clients were asked each of the six questions either verbally on the phone, or in written form by mail. The verbal contact allowed for more accuracy in answers since the interviewer could clarify the information that was being provided. Ultimately, the client responses had to be interpreted by the interviewer, and the data collected was coded based on the subjective interpretation of the interviewer. The mailed responses also required some interpretation since the answers were of varying lengths and specificity. The interpretation was based on previous assessments of the client by the providing counselor.

Results

Sample characteristics (Table 2 and 3). For intake and discharge data, 66% of the sample was male and 34% female. The mean age for the sample was 38, for males 37, and females 39. The sample was divided among two client populations with 38% clinic and 62% DUII. The average days in treatment for the sample was 154.2, for clinic clients 159 and for DUII's 151. The average days in treatment for a male was 147.2 compared to 167.6 for females. A summary of results from the SASSI, BDI and Folstein can be found in Table 2.

Of the twenty-nine clients enrolled in the study, only twenty-two (76%) were reachable and provided data for the study. The follow-up sample included 72% males and 28% females. It was divided among two client populations with 28% clinic and 72% DUII. The goal was to

contact clients no less than 30 days after discharge, the average follow-up data was collected 39 days following discharge.

ASAM dimension question results. Outcome data was collected on all six dimensions with details of results in Table 4, 5, and 6. A review of the results for the entire sample are included here.

Dimension One. The mean alcohol use in the last 30 days was compared at intake (7.5 days), discharge (2.2 days), and follow-up (2.9 days). There was a significant decrease in consumption from intake to discharge ($t(df=28) = 3.513, p < .01$), and from intake to follow-up ($t(df=21) = 2.574, p < .01$), but not between discharge and follow-up. Mean drug use in the last 30 days was similarly compared, with use at intake (2.7 days), discharge (1.1 days), and follow-up (.9 days) yielding no significant changes. The percent of the sample population responsible for alcohol consumption was also compared at intake (45%), discharge (14%), and follow-up (55%). There was a significant reduction in the percentage of clients who used alcohol at discharge as compared to intake (McNemar $\chi^2 = 8.1, df = 1, p < .01$), but a significant increase in the percentage of those who used alcohol between discharge and follow-up (McNemar $\chi^2 = 11.53, df = 1, p < .01$). There was no significant changes from intake to follow-up. The percent of the sample population responsible for drug use was also compared at intake (21%), discharge (10%), and follow-up (23%). No significant changes were found, but change reflected the results found with alcohol in that there was a decrease from intake to discharge, and an increase from discharge to follow-up.

Dimension Two. The percent who reported physical health issues or problems were compared at intake (45%), discharge (21%), and follow-up (32%). There was a significant reduction in the percentage of clients reporting health problems at intake as compared to discharge (McNemar $\chi^2 = 5.14, df = 1, p < .05$), but no significant changes between discharge and follow-up, or between intake and follow-up.

Dimension Three. The percent who reported emotional/mental health problems were compared at intake (34%), discharge (31%), and follow-up (23%). Chi-square analysis showed no significant changes, but results indicate a decrease in symptomatology over time.

Dimension Four. Results from question four indicate that acceptance of life issues or problems is high at intake (90%), increases throughout therapy until discharge (97%), and then

decreases at follow-up (77%). Chi-square analysis showed no significant changes, but the change between discharge and follow-up showed a trend toward decrease in acceptance.

Dimension Five. Results from question five indicate that the percentage of persons in each relapse category (low, medium, high) changes significantly from intake to discharge ($\chi^2 = 8.643$, $df = 1$, $p < .05$). Results indicate that at intake 19% were rated as a low risk for relapse, 38% moderate, and 43% as high. At discharge this had changed to 47% low, 44% moderate, and 9% high. The percentage of persons in each category at follow-up were 36% low, 56% moderate and 28% high. Although the percentage of persons falling into each category at follow-up is different than at intake and/or discharge, these changes were not significant.

Dimension Six. Question six assessed the percentage of persons noting a change in their recovery environment at intake (24%), discharge (41%), and follow-up (27%). Chi-square analysis yielded no significant changes, but the pattern of increase between intake and discharge, and decrease between discharge and follow-up found in questions 2, 4, and 5 was also seen in this case.

Discussion

Results from dimension one suggest that alcohol and drug consumption at discharge and follow-up are below levels at intake. However, the percent of the population that used alcohol/drugs remained fairly constant, therefore the decrease in consumption appears to be linked to a portion of the sample. Since this study did not differentiate between abusing and dependent clients, conclusions regarding which of these types of clients are responsible for the improvement cannot be made. DUII clients showed an increase in alcohol consumption at follow-up probably due to no longer having to remain abstinent as part of a treatment requirement. In addition, the increase in percent of the DUII population responsible for drug use appears to be linked to occasional cannabis use upon reviewing follow-up data. It is important to note that since this study did not utilize a true experimental design with a control group, concluding that the observed changes were a direct result of treatment is not possible. The outcome results could be attributed to other variables including chance events, client attributes, or problems in our measurement procedures. This point holds true for the data collected on the remaining questions as well.

Results from dimension two suggest that physical health problems reported at intake are significantly reduced during the time clients are in treatment. Since counselors at BHC are trained to address issues using a biopsychosocial model, clients with physical health problems are encouraged to seek medical attention and follow-up on physical health issues. Since research has shown a strong correlation between mental health/addictions and physical health problems (Haber & Mitchell, 1997), it can be hypothesized that clients physical health improved as they begin to work on their behavioral health issues. For example, learning relaxation techniques in group may help to reduce stress, which in turn helps to lower a client's blood pressure. Their health also improved in treatment because they were encouraged to seek medical treatment if needed.

Although changes in mental or emotional health were not significant, they did show improvement over time. An interesting finding with clinic clients was that improvement was observed between intake and discharge, but increased significantly at follow-up. One possibility is that once out of treatment, clients tended to continue to use the tools learned in treatment to improve their life situation. Alternatively, it is also possible that once away from treatment, clients thought less about having a mental health problem since it was no longer reinforced by attending a treatment clinic, and therefore reported less symptoms at follow-up. DVI clients mental health issues decreased at discharge, but increased at follow-up. Since DVI clients are mandated to be in treatment, the initial improvement may be a result of clients who receive some benefit from having emotional issues discussed in treatment, but return to old patterns of behavior once they leave the program. It is also possible that DVI treatment may uncover longstanding emotional issues that have remained unaddressed, but once identified in treatment become ongoing issues for clients once they complete the DVI program. Although BHC allows clients to continue at the clinic even after DVI treatment completion, reviewing the core data set it appears that only a small percentage actually stay in treatment.

The results for question four show acceptance increasing at discharge, but then significantly declining at follow-up. It can be hypothesized that treatment helps clients gain a deeper level of acceptance of their life issues through ongoing evaluations, assessments, and interaction with other group members. However, it is also possible that counselors misinterpret clients' reactions and feelings about acceptance of treatment, or that clients are not entirely truthful about their feelings regarding treatment. The decline at follow-up may be explained by

The purpose of the pilot study was twofold; to begin an ongoing evaluation at an outpatient additions clinic, and show how the fundamental evaluation steps are applied in practice. The study also provided feedback regarding implementation issues that arose during the project. Since no study had ever been conducted at BHC, the pilot study was an attempt to begin collecting a core data set, and use additional information that was currently being collected to assess overall program effectiveness. The pilot study was the first step in designing an on-going evaluation system at BHC. It provided an opportunity to implement the evaluation steps to a

Conclusions

set-backs in treatment, relapse behavior, or problems related to treatment techniques and/or counselor attributes. It is important to note that the high acceptance rates on DUII clients were due to acceptance of driving while intoxicated and being required to complete a DUII program, and not necessarily due to admission of alcohol/drugs problems or needing treatment. This in part explains the drop in acceptance at follow-up. Since the pilot study did not use any standardized assessment measures, making conclusions, especially on constructs such as acceptance, need to be viewed with significant caution.

Relapse risk appears to decrease significantly from intake to discharge, and then increase at follow-up. The suggestion is that being in treatment reduces relapse risk, and without the support of treatment relapse risk increases. It appears that once clients leave treatment, many do not follow through on their recovery program goals. This lack of follow-up was interpreted by research evaluators as behavior associated with higher risk for relapse. However, it is also probable that many clients may procrastinate recovery work once out of treatment until issues resurface that require attention. Unfortunately, the design of this study did not allow for any conclusion about whether treatment was the variable responsible for the change, only that change did occur.

It appears that treatment may also impact client's recovery environment. Changes reported over time include moving, securing new friends, joining new support groups, and switching jobs. The greatest level of change was reported at discharge, but whether these changes can be attributed to treatment cannot be determined. It does point out that in addition to changes in substance use and mental health issues, many clients are also experiencing changes in their environment that need to be addressed in treatment.

sample of the clinic's population, and determine what evaluation changes need to be made before evaluating a larger subgroup, or the entire population. In reviewing the fundamental research steps, and evaluating the pilot project, it is apparent that there are a number of conclusions that can be made.

The pilot study strayed at times from the fundamental evaluation steps, and as a result provided results with limited specificity and objectivity. It attempted to collect both summative and formative data, and answer six different evaluation questions instead of carefully crafting one question. In addition, the pilot study utilized a pretest/posttest design which did measure change, but did not specify what variables were responsible for the change. Instead of selecting a standardized measurement instrument to assess change over time, an assessment questionnaire was created that had untested reliability and validity. The answers collected from the questions asked at follow-up involved subjective interpretations by the researcher and therefore must be considered as preliminary outcomes, that should be explored in more depth in future evaluations.

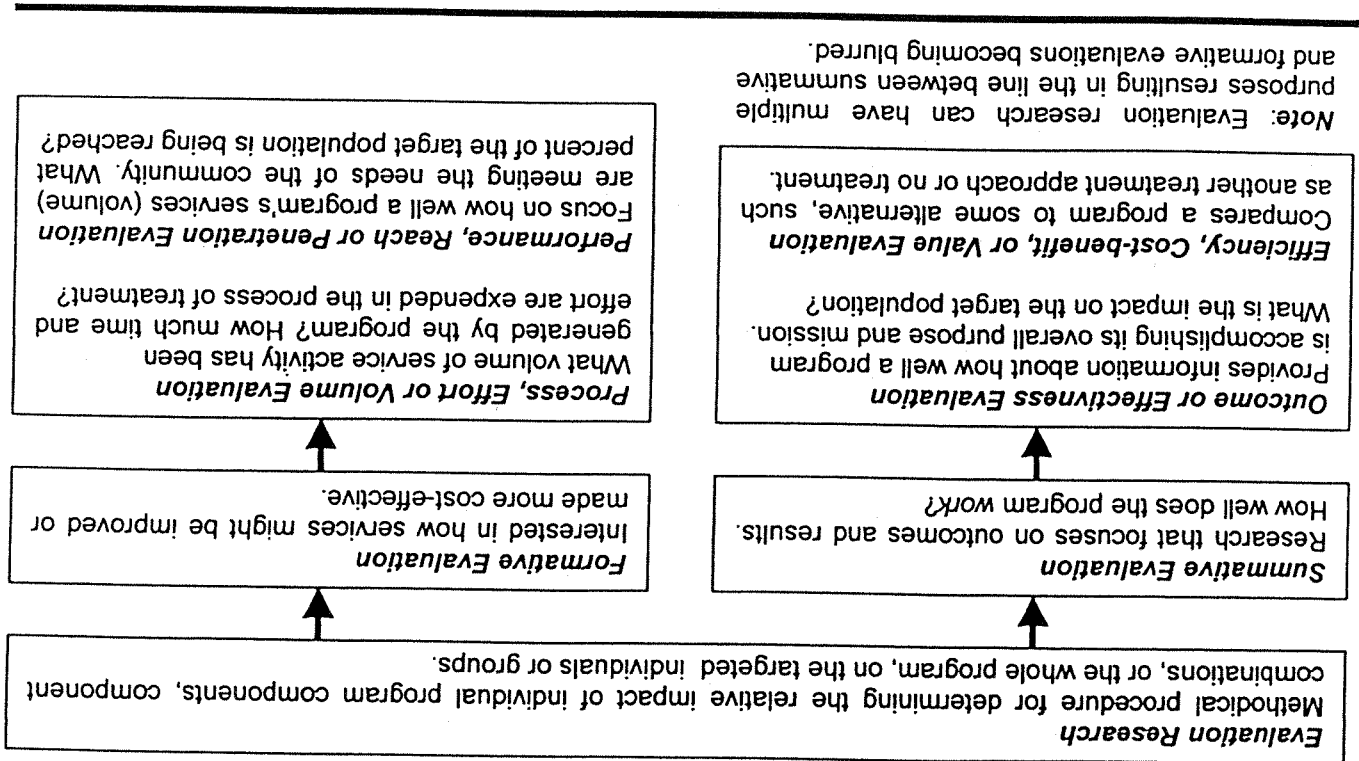
Recommendations

It is suggested that the clinic view the pilot study as the beginning of an on-going evaluation process to help improve overall delivered services. Ideally, evaluation research would become part of the culture at BHC, and be integrated into all aspects how the clinic operates and services clients. Since the pilot study revealed the difficulties in asking multiple evaluation questions, the next evaluation should focus on one or two specific research questions that can be addressed in sufficient depth to provide accurate and useful information. It is advisable that standardized measurement instruments be used to increase the validity and reliability of the future evaluations at BHC can be made.

Following the fundamental evaluation steps does not necessarily mean evaluation research will always be easy. In the BHC pilot study all staff showed initial support for the project and a willingness to participate in the process. However, as the study developed and participants were needed, support waned due primarily to counselors having limited time or energy to devote to the project. With the burden falling primarily on one person who was also working full-time as a counselor, it became apparent that conceptually, conducting evaluation research makes a lot of sense; but finding the time, resources and energy to implement a study alone can be challenging. From this pilot study, a number of recommendations for improving

results, instead of creating a measure such as the questionnaire used in the pilot study. Additionally, due to the previously discussed limitations in the pretest/posttest design of the pilot study, future evaluations should consider using an experimental, or quasi-experimental design to determine what variables are responsible for noted changes. Although applying this type of design to the entire population could be problematic, it may be possible to utilize one of these designs in a specific subpopulation or group. Additionally, since data collection did include collecting scores on standardized instruments (BDI, SASSI, Folstein) at intake, for future evaluations it would be advisable to retest clients at discharge and follow-up with the same instruments, significantly improving the overall reliability and validity of the evaluation. Finally, it is advisable if a clinic is serious about evaluation research, and has management support, that all counselors need to have some involvement throughout the process. They are the agents delivering treatment to clients on a day-to-day basis, and need to be comfortable with the evaluation process and how the results will be used.

Figure 1
Classification of evaluation research types



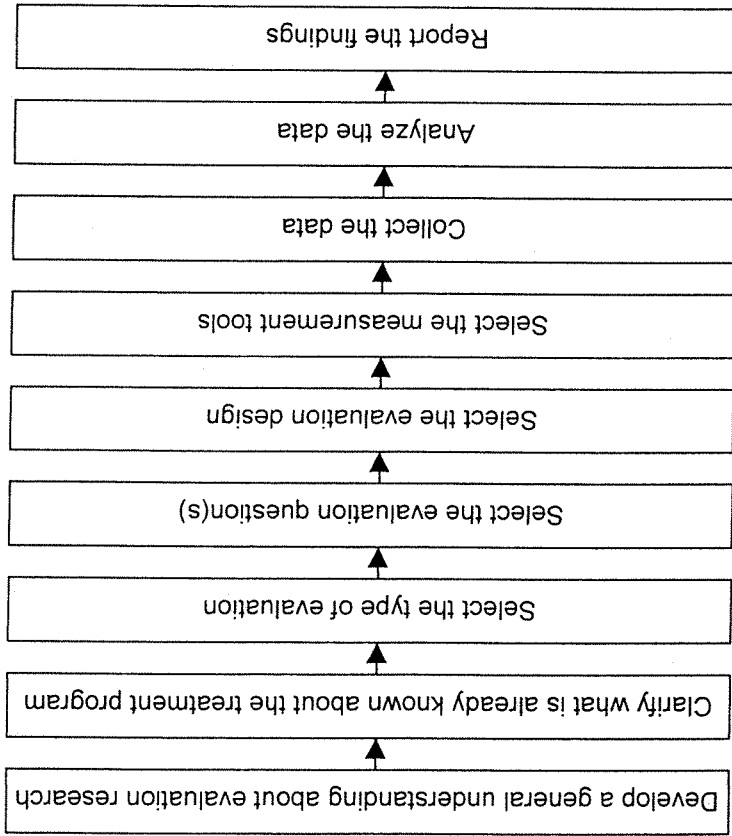


Figure 2
Research evaluation steps

Table 1
Core data set samples

Yearly Clients	
Total clients from 10/1/97 to 9/30/98	717
Total clinic clients	458
Total DUII clients	186
Total gambling clients	73
Percent female client population	25%
Percent male client population	75%
Revenues	
Client payment	11%
Private insurance	1%
Open card	1%
Managed care	12%
Medicare	1%
State contract	28%
Gambling contract	46%
Monthly Report	
Enrolled clients	134
Gender	
Male	102 (76%)
Female	32 (24%)
Race	
White	101
Black	10
Hispanic	2
Asian	7
A. Indian	4
Unknown	10
Other	
Alcohol only	43%
Drug only	18%
Both alcohol and drug	39%
IV drug use at admit	10%
Known to have TB	<1%
Known to have HIV	<1%
Previous treatment	75%
Pregnant at intake	<1%

Table 3
Demographic results at follow-up

	N	Percent	Male	Female
Clinic	6	28%	3	3
DUII	16	72%	13	3
All Types	22	100%	16	6

Table 5 Results of changes in alcohol and drug consumption at intake, discharge and follow-up

	t-test
Alcohol Consumption: Intake/Discharge	3.513 **
Alcohol Consumption: Discharge/Follow-up	-.548
Alcohol Consumption: Intake/Follow-up	2.574 **
Drug Use: Intake/Discharge	2.032
Drug Use: Discharge/Follow-up	.755
Drug Use: Intake/Follow-up	1.183

** Significant at p<.01

Table 6 Results of changes in ASAM dimensions at intake, discharge and follow-up

	McNemar Chi
Alcohol Consumers: Intake/Discharge	8.1 **
Alcohol Consumers: Discharge/Follow-up	11.53 **
Alcohol Consumers: Intake/Follow-up	1.45
Drug Users: Intake/Discharge	1.33
Drug Users: Discharge/Follow-up	1.33
Drug Users: Intake/Follow-up	.25
Physical Health: Intake/Discharge	5.14 *
Physical Health: Discharge/Follow-up	.5
Physical Health: Intake/Follow-up	1.33
Mental Health: Intake/Discharge	0
Mental Health: Discharge/Follow-up	0
Mental Health: Intake/Follow-up	0
Acceptance: Intake/Discharge	.5
Acceptance: Discharge/Follow-up	2.25
Acceptance: Intake/Follow-up	.17
Relapse Risk: Intake/Discharge	8.64 *
Relapse Risk: Discharge/Follow-up	2.52
Relapse Risk: Intake/Follow-up	1.83
Environment: Intake/Discharge	2.29
Environment: Discharge/Follow-up	0
Environment: Intake/Follow-up	.57

** Significant at p<.01
* Significant at p<.05

References

- Carnes, P. (1989). *Contrary to love: Helping the sexual addict*. Center City, MN: Hazelden Educational Materials.
- Dornelas, E.A., Correll, R.E., Lothstein, L., Wilbur, C., & Goethe, J.W. (1996). Designing and implementing outcome evaluations: Some guidelines for practitioners. *Psychotherapy, Vol. 33, Summer, No. 2, 237-245*.
- Eisen, S.V. & Dickey, B. (1996). Mental health outcome assessment: The new agenda. *Psychotherapy, Vol. 33, Summer, No. 2, 181-189*.
- Graham, K. (1994). Guidelines for using standardized outcome measures following addiction treatment. *Evaluation & The Health Professions, Vol. 17, No. 1, March, 43-49*.
- Haber, J.D. & Mitchell, G.E. (Eds.). (1997). *Primary care meets mental health: Tools for the 21st century*. Tiburon, CA: Centralink Publications.
- Johnson, L.D. (1995). *Psychotherapy in the age of accountability*. New York, W.W. Norton.
- Johnson, L.D. & Shaha, S. (1996). Improving quality in psychotherapy. *Psychotherapy, 33, 225-236*.
- Kiesler, C.A., Simpkins, C.G., & Morton, T.L. (1991). Prevalence of dual diagnosis of mental and substance abuse disorders in general hospitals. *Hospital and Community Psychiatry, 42, 400-403*.
- McLellan, A.T. & Durell, J. (1996). Outcome evaluation in psychiatric and substance abuse treatments: Concepts, rationale, and methods. In L.I. Sederer and B. Dickey (Eds.), *Outcome Assessment in Clinical Practice* (pp. 34-44). Baltimore, MD: Williams and Wilkins.
- Miller, W.R. (1997). Why do people change addictive behavior? The 1996 H. David Archibald Lecture. *Addiction, 93 (2), 163-172*.
- Miller, W.R. & Brown, S.A. (1997). Why psychologists should treat alcohol and drug problems. *American Psychologist, 52, 1269-1279*.
- Miller, W.R. & Wolfe, B.L. (1996). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: Dept. of Psychology, U. of New Mexico.
- National Institute of Health. (1993). *How good is your drug abuse treatment program?* Rockville, MD: National Institute of Health.

- Pheiffer, S.I., Soldivera, S., & Norton, J.A. (1992). A consumer's guide to mental health outcome measures. Devon, PA: The Devereux Foundation.
- Prochaska, J.O., Norcross, J., & DiClemente, C.C. (1994). Changing for good. New York, NY: Avon.
- Rainer, J.P. (1996). The pragmatic relevance and methodological concerns of psychotherapy outcome research related to cost effectiveness and cost-offset in the emerging health care environment. Psychotherapy, 33, 216-224.
- Regier, D.A., Farmer, M.F., Rae, D.S., Locke, B.Z., Keith, S.J., Judd, L.L., & Goodwin, F.K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse. Journal of the American Medical Association, 264, 2511-2518.
- Schlosser, B. (1995). The ecology of assessment: A "patient-centric" perspective. Behavioral Healthcare Tomorrow, May/June, 66-88.
- Shaffer, H.J. (1996). Understanding the means and objects of addiction: Technology, the internet, and gambling. Journal of Gambling Studies, 12, 461-469.
- U.S. Department of Health and Human Services. (1987). Handbook for Evaluating Drug and Alcohol Prevention Programs. Rockville, MD: U.S. Department of Health and Human Services.
- U.S. Department of Health and Human Services. (1995). Developing State Outcomes Monitoring Systems for Alcohol and Other Drug Abuse Treatment, TIP series 14. Rockville, MD: U.S. Department of Health and Human Services.