Workbook 7

Outcome Evaluations

WHO
World Health Organization

UNDCP
United Nations International Drug Control Programme

EMCDDA
European Monitoring Center on Drugs and Drug Addiction

Acknowledgements

The World Health Organization gratefully acknowledges the contributions of the numerous individuals involved in the preparation of this workbook series, including the experts who provided useful comments during its development for the Substance Abuse Department, directed by Dr. Mary Jansen. Financial assistance was provided by UNDCP/EMCDDA/Swiss Federal Office for Public Health. Christine Godfrey (United Kingdom) wrote the original text for this workbook, and Brian Rush (Canada) edited the workbook series in earlier stages. JoAnne Epping-Jordan (Switzerland) wrote further text modifications and edited the workbook series in later stages. Munira Lalji (WHO, Substance Abuse Department) and Jennifer Hillebrand (WHO, Substance Abuse Department) also edited the workbook series in later stages. This workbook's case examples were written by Michael French and Kerry Anne McGeary (U.S.A.); by Ellen Williams and Dean Gerstein (U.S.A.), and were edited by Ginette Goulet (Canada). Maristela Monteiro (WHO, Substance Abuse Department) provided editorial input throughout the development of this workbook series.

Table of contents

Overview of workbook series

This workbook is part of a series intended to educate programme planners, managers, staff and other decision-makers about the evaluation of services and systems for the treatment of psychoactive substance use disorders. The objective of this series is to enhance their capacity for carrying out evaluation activities. The broader goal of the workbooks is to enhance treatment efficiency and cost-effectiveness using the information that comes from these evaluation activities.

This workbook considers outcome evaluation. Outcome evaluations measure the extent to which clients of services, or networks of services for substance use disorders, change following participation in treatment. The workbook offers advice on measuring the changes and attributing change to programme involvement.

Introductory Workbook
Framework Workbook

Foundation Workbooks
Workbook 1: Planning Evaluations
Workbook 2: Implementing Evaluations

Specialised Workbooks
Workbook 3: Needs Assessment Evaluations
Workbook 4: Process Evaluations
Workbook 5: Cost Evaluations
Workbook 6: Client Satisfaction Evaluations
Workbook 7: Outcome Evaluations
Workbook 8: Economic Evaluations

What is an outcome evaluation?

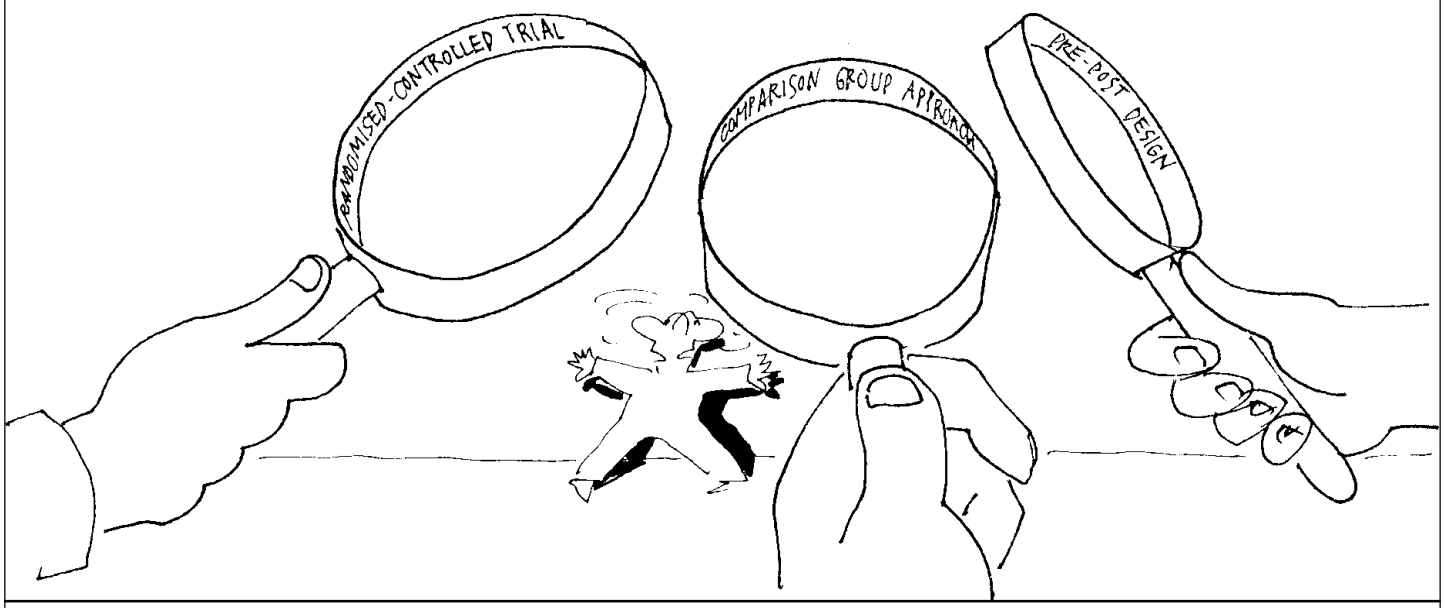
Outcome evaluations provide information on how well your programme is accomplishing its goals.

Outcome evaluations **measure how clients and their circumstances change**, and **whether the treatment experience has been a factor in causing this change**. In other words, outcome evaluations aim to assess **treatment effectiveness**.

Some questions that might be addressed in outcome evaluations include:

- Have clients' quality of life improved following treatment?
- Has there been a reduction in the quantity/frequency of PSU following treatment?
- Is client participation in our treatment programme "responsible" for their improvement?

There are a number of ways to design outcome evaluation and measure these types of changes. The most widely-praised way to measure client improvement and infer causality (i.e., to infer that your programme is responsible for the observed client improvement) is the experimental approach. This is sometimes called a "randomised-controlled trial". Other methods for studying outcome include the "comparison group approach" and the "pre-post design." All these methods are described later in this workbook.



Why do an outcome evaluation?

If clients get better following treatment, it does not necessarily mean that your treatment caused these changes.

If your programme is typical, several groups of people would be interested to know whether your treatments are actually effective.

- Your clients
- Their family members
- Your treatment staff
- Employers of clients
- Criminal justice system
- Health insurers or other "payers" for your treatment services
- Government organisations
- General community members

You may be saying to yourself, "I know that my treatment programme is effective because I have seen many people get better following participation. Why do I need to do an outcome evaluation?" The answer is relatively simple. If clients get better following treatment, it does not necessarily mean that your treatment **caused** these changes. Think about it. People change for many reasons. Improvements in your clients' PSU may be the result of something completely different from your programme. Common other reasons for improvement, beyond the effects of treatment itself, include:

• Other things that happened during and after treatment

Clients may have found or lost an important interpersonal relationship; found or lost a job; moved to a new neighbourhood; or become involved with a self-help group. All of these events could influence their PSU independently of the effects of treatment. Such events can also interact with treatment effects in complex ways. For example, the clients who have done well may be those who experienced other positive life events.

• Client "maturation"

Over time, many people grow out of their problems due to age-related changes. This is particularly so for adolescents whose PSU tends to decrease as they reach young adulthood.

• Natural variation or regression

Although many clients of services for PSU disorders, lead disruptive lives, they also have periods of relative stability when they cut down or eliminate their PSU. Any changes in the behaviour and circumstances between two periods of time may simply reflect "normal" variations rather than the effects of an intervention. Some of those who enter treatment during a particularly disruptive period can be expected to change for the better without treatment, even if temporarily.

For these reasons, outcome studies go beyond merely describing positive changes in clients. They attempt to demonstrate scientifically whether your treatment process has caused any client changes that occur.

How to do an outcome evaluation?

Your choice of designs should be influenced by the resources you have available.

Outcome evaluation is based on a quantitative approach. It typically uses one of three designs:

- randomised controlled trial
- comparison group
- pre-post comparison

Each of these designs are described below. After reading this workbook, you must make your choice among these design options. In general, pre-post comparison is the least scientifically rigorous design, comparison group designs are "moderate" in their scientific rigour, and randomised controlled trials use the strongest design. However, randomised controlled trials and comparison group designs are more resource-intensive and complicated to conduct than pre-post comparisons. Your choice of designs should be influenced by the resources you have available. After reading this material, you must carefully consider the practical realities of implementing each type of design in your programme setting.

Method 1: Randomised controlled trial tesign

...clients are randomly assigned (like the flip of a coin) to either the treatment in question or to a plausible alternative.

This design option uses two or more groups of clients who are **randomly assigned** to either the treatment in question or to a plausible alternative. Members of both groups receive the same pre-treatment and post-treatment assessments. Because the randomisation process makes it equally likely that any one client will be assigned to one group or the other, with a sufficient number of participants this design controls for pre-treatment individual differences in clients (e.g., PSU frequency, motivation for treatment) and other events that might happen during treatment.

Randomised controlled trials can compare many things, including different types of treatment, (e.g., pharmacotherapy vs. psychotherapy); different intensities of the same treatment (e.g. short vs. long-term); different strategies for delivering the same treatment (e.g., group vs. individual); and different settings (e.g., inpatient vs. outpatient). Other comparisons involve people who received no treatment vs. people who receive treatment.

A significant strength of randomised controlled trials is that they can control for most competing explanations for improvement following treatment (e.g., other events that happened during treatment). However, there are many technical and logistical problems to overcome in the proper design and conduct of these evaluations. Consultation with an evaluator experienced with randomised controlled trials is recommended if you are considering this design.

The first case report located at the end of this workbook (by Formigoni and Marques) provides an example of a randomised controlled trial design. In this evaluation, individual and group cognitive-behavioural treatments were compared using random assignment to treatment conditions.

Method 2: Comparison group designs

The success of the evaluation depends on how similar the two groups are at the beginning of the evaluation.

This design option is similar to the randomised design except the comparison group is deliberately rather than randomly chosen. Comparison groups are chosen so that clients are as similar as possible to those in the treatment service or system being evaluated. Statistical methods are used to control for any remaining differences (e.g., differences in client age). The types of treatment and alternative conditions featured in comparison group evaluations are similar to those noted above in connection with experimental evaluations (i.e., comparisons treatment types or intensity).

The extent to which comparison group studies successfully control for the various competing explanation factors varies with the types of services and client groups involved. The "success" of the evaluation depends on how similar the two groups are at the beginning of the evaluation. For example, different types of clients may have different reasons for their choice of treatment programme. Similar to randomised controlled trials, there are many technical and logistical problems to overcome in the proper design and conduct of these evaluations. Consultation with an evaluator experienced in comparison group evaluations is recommended if you are considering this design.

Method 3: Pre-post design

Pre-post studies assess clients on the same variables, and over the same time intervals, before and after they complete treatment

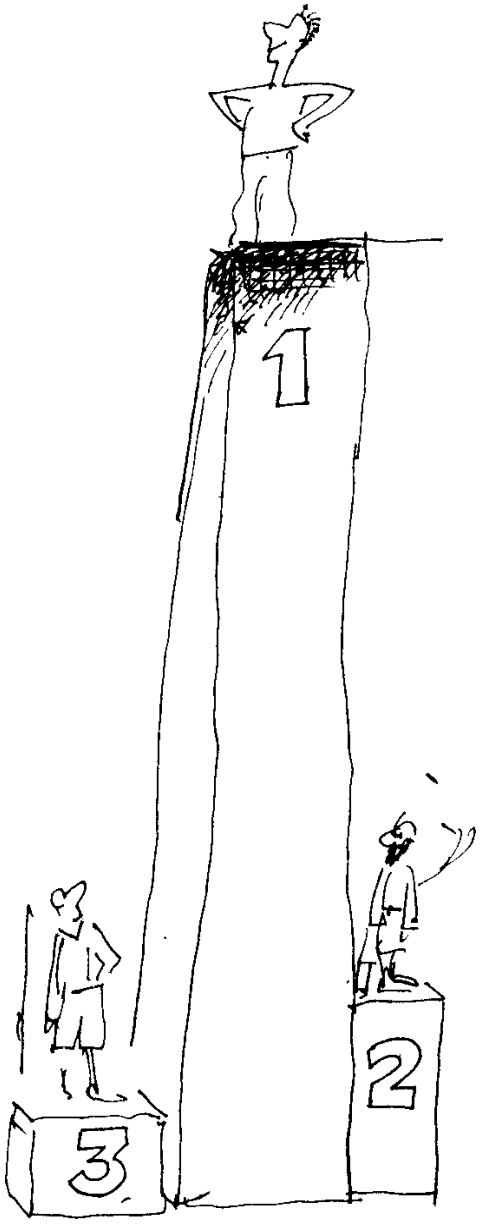
This design option is not as complex as experimental and comparison group evaluations. It is more realistic for treatment services or systems with limited experience and/or resources. Although pre-post designs are less scientifically rigorous, they can produce useful results for purposes of accountability and programme improvement.

Pre-post studies assess clients on the same variables, and over the same time intervals, before and after they complete treatment. For example, baseline data collected at intake may ask about quantity and frequency of PSU over the past 90 days. Exactly the same questions would be asked of clients six months following discharge. With a pre-post design, clients may be re-contacted on more than one occasion (e.g., six months, 12 months, 18 months, and 24 months). In this case, the evaluation is called a time-series design. While this extended design is especially helpful in showing the stability of outcomes being achieved, an extra effort is required to maintain contact with the sample of clients being followed-up.

Simple pre-post designs have some limitations. They do not control for competing explanations, such as something else that happened during treatment. They also do not show if a treatment service or system is effective **relative to alternative treatments**.

Despite these drawbacks, pre-post evaluations have several strengths. They can determine if treatment objectives are being achieved, and the type of client who improves most or least. They also can show if improvement varies with the amount or type of treatment received. For example, you can show if those attending treatment consistently fare better than those with poor attendance. You also can show if those who attended specific components of the service did better than others. Positive results from a pre-post study are rewarding for staff and can satisfy accountability requirements of some funding agencies. Positive results from pre-post studies also can lend support for getting the resources to conduct more rigorous randomised controlled trials. If your results seem worse than those reported for similar clients in other programmes, some action to improve your services may be needed. If, on the other hand, the results seem better than expected, a rigorous evaluation may be desirable to be sure your programme can take the credit for these positive outcomes.

The last case report located at the end of this workbook (by Auriacombe and colleagues) provides an example of pre-post design. In this evaluation, clients receiving opiate-substitute treatment were followed over time.



Other methodological issues in outcome evaluation

Beyond choosing the basic design of your evaluation, there are other methodological details to be decided. Workbooks 1 and 2 provide valuable general information in this area; be sure to review them with this workbook. The information in this section is specialised for outcome evaluations and complementary to the more general information provided in the introductory workbooks. The following issues are discussed:

1 selecting clients for participation

2 sample size

3 timing and frequency of follow-up

4 preparing and tracing clients for follow-up interviews

5 conduct of follow-up interviews

6 selection and training of interviewers

1. Selecting clients for participation

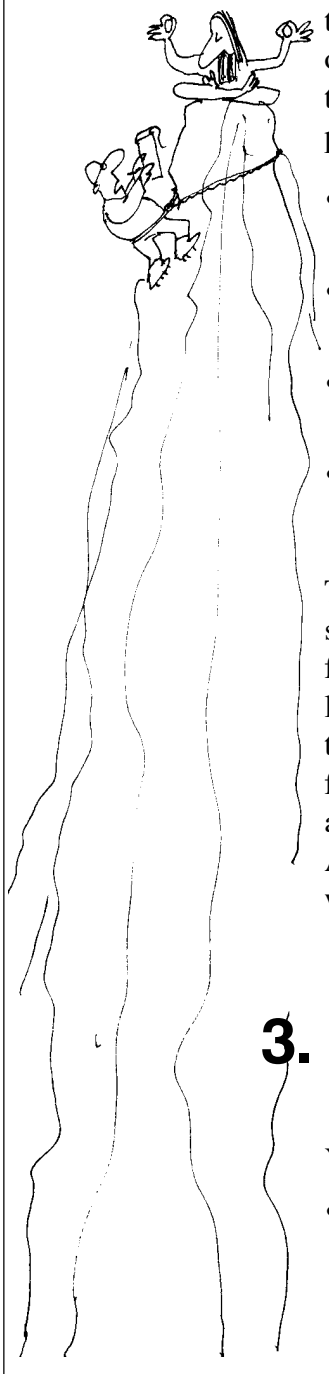
The selection of clients for participation in an outcome evaluation should be determined by objectives of the evaluation. If you are interested in general programme effectiveness, random samples of **all** clients who enter treatment in a typical time period should be selected.[1] If, on the other hand, your objectives of evaluation concern particular types of clients (e.g., opiate users), or clients who complete a certain amount of treatment, then random samples should be chosen to represent this subgroup. If there is a desire to compare one service with another, then similar types of cases should be recruited from each service.

Beyond taking a random sample, there are no stead-fast rules about who to enrol. However, the procedures used to select clients should be stated clearly to ensure individuals who read your evaluation report understand the procedures and potential biases. Attention should be drawn to clients who were excluded from participation, such as clients who don't have a telephone, because exclusions can affect your results.

[1]One could argue that all clients, as opposed to a sample of clients, should be routinely followed up for purposes of accountability. This is not usually feasible given the time and resources required to do so.

2. Sample size

More cases are required if you want to detect smaller differences between groups.



There are no simple answers to the question, "How many clients do I need to study?" Much depends on the objectives of the study, the kinds of clients involved and the kinds of measures used. Case examples from this workbook report evaluations with as few as 16 participants to greater than 1,000 participants. If the aim is to compare outcomes of two groups of clients (i.e., males or females; two programs), the number to be studied depends on the size of difference you want to detect between the two groups on the outcome measure. A statistician will be able to calculate the required sample size if you provide the following information:

- the relative sizes of the two groups to be compared
- the expected frequency of the behaviour in one group
- the magnitude of the difference that you want to be able to detect, between groups
- the degree of confidence you want to have in the results

To show you how this process works, consider this example. Imagine that you want to find out if males in your programme are more likely to relapse within the first three months than females. Assume that you will have data for an equal number of males and females and that you expect 40% of males to relapse. Assume further that a difference of 20% would be of practical significance, and that you want to be 95% certain that any such observed differences were not due to chance. In this case, the statistician will likely advise you to collect data on about 180 cases (90 males and 90 females). If, however, you think that a difference of 10% between males and females is likely to be of interest, the statistician will advise you to collect data on 600 cases. More cases are required if you want to detect smaller differences between groups. More cases are also required if more than two groups are to be compared or if the groups are of unequal size.

When planning the number of people to be studied in an outcome evaluation, allowance should be made for clients who cannot be contacted and for whom outcome information will be missing. Remember, your final sample calculations will be based on the number of clients for whom you have complete data. You will have to contact more clients in order to get this many for final calculations. The percentage of clients "lost" to follow-up will vary from situation to situation. It will depend to some extent on the social stability of clients and the ingenuity of follow-up workers. It would be reasonable to expect that up to 30% of cases chosen for follow-up cannot be traced and to, therefore, increase the sample selected for follow-up by 30%. McLellan and colleagues (1996) recommend a 70% follow-up rate as the minimum standard for outcome evaluation.

3. Timing and frequency of follow-up

You have three factors to consider here:

- the point in time at which you start counting weeks/months until the follow-up interval (i.e., 4 weeks after intake and assessment vs. 4 weeks after some period of treatment participation vs. 4 weeks after the last treatment contact or formal discharge)
- the duration of the follow-up interval (e.g., 4 weeks vs. 8 weeks vs. 12 weeks)
- the time period over which PSU and other outcomes are assessed

Start Date

...we recommend that you start the follow-up period at the first face-to-face contact for client assessment.

In selecting the start date for the follow-up period, there are several trade-offs to be made. If you decide to follow-up a random sample of clients who are enrolled at intake and/or assessment, you will obtain the largest sample. Using this method, results can be generalised to all clients who have participated in the programme regardless of the level of service they eventually receive. On the other hand, most clients who drop-out of treatment do so early in the treatment process; many after their first contact. Selecting your sample this early in the process will mean more effort to locate people for follow-up, because early drop-outs will be more difficult to locate.

If you select your follow-up sample from those who complete a certain period of treatment, or who have made a certain number of contacts (e.g., three outpatient visits), you will have a more stable group to re-contact. You will, however, have missed the opportunity to determine outcome for those with fewer contacts. If you contact only those completing treatment, and who are formally discharged, you will probably have a sample heavily biased toward positive outcome.

Given the above considerations, we recommend that you start the follow-up period at the **first** face-to-face contact for client assessment. McLellan and colleagues (1996) refer to this as the "intent-to-treat" design and recommend it as a minimum standard for outcome evaluation.[2] With this approach, your baseline evaluation information must be collected as early as possible in the intake/assessment process. In many programmes, clinical assessment and treatment planning extends over several contacts. For the period of the evaluation, routine assessment procedures may need to be modified in order to get the pre-treatment evaluation information at the first contact for assessment.

The relationship between standard programme intake/assessment procedures and the collection of the pre-treatment evaluation data requires careful planning. One option is to conduct an evaluation interview **in addition to** the normal assessment protocol. The interview may be conducted by an independent evaluator or by other programme staff. The disadvantage of this approach is that the client may be overburdened by two data collection procedures that ask for similar information in slightly different ways. An alternative is to blend the pre-treatment evaluation questions into the clinical assessment process. This has the advantage of reducing the burden on clients and maximising the use of staff resources if independent evaluators can not be used.

[2]Many programmes have a clerical function incorporated into the initial stage of treatment involvement which collects basic demographic information and screens the client for programme eligibility. It is very difficult to collect baseline evaluation data during such a contact with the programme. Thus, the "intention-to-treat" design often means that clients are selected for evaluation at the point of their first **clinical encounter** for assessment and/or treatment.

Duration of follow-up period

It iscommended that you select at least a six-month follow-up interval and consider the potential benefits of at least one additional contact in another few months.

Your second major decision is the duration of the follow-up interval. Follow-up studies published in research journals have reported on information obtained at many different points in time after treatment engagement. Some evaluations report on client changes at the time of discharge or after a certain period of outpatient contact. Other evaluations have followed a sample of clients for several years. Most common are reports of outcomes assessed after a three, six, or 12 month interval. Outcome studies to be reported in scientific journals typically require a one to two year follow-up period. The case examples at the end of the workbook demonstrate this variability. Two evaluations used a 12-month follow-up, whereas the third followed participants for five years.

The timing of your follow-up will have a significant impact on your results and conclusions. Short-term follow-up studies will show better results than longer term ones, because 60%-80% of "relapses" occur in the first three to four months following discharge (McClellan et al., 1993).

It is recommended that you select at least a six-month follow-up interval and consider the potential benefits of at least one additional contact in another few months. However, there are no hard and fast rules that must be followed, other than that your follow-up periods and intervals should be consistent with the objectives of your evaluation.

Time period for measures

...it is recommended that you select a 90-day period for your outcome measures.

Your third major decision concerns the time period over which outcomes will be assessed. For example, even though you may have decided that your follow-up period will be six months in duration, you still need to decide the time period over which clients will be asked to recall their PSU and its consequences. The same time period must be chosen for both the pre-treatment and post-treatment assessments.

There are trade-offs for any time period you choose. A client's PSU in the 30 days prior to starting treatment may not be representative of longer term PSU. Thus, comparison of the 30-day pre-treatment period and a 30-day post-treatment period may not yield a reliable and meaningful difference. On the other hand, if the time period is too long (e.g., 4-6 months), clients may not be able to recall important information accurately (e.g., frequency and quantity of PSU; use of health and correctional services).

Based on these concerns, it is recommended that you select a 90-day period for your outcome measures. This time period will need to be stated clearly to clients and reflected in your questionnaires during pre-treatment and post-treatment assessments.

4. Preparing and tracing clients for follow-up interviews

The consent form should indicate the reason you are evaluating clients, the (random) process of selection, assurances of confidentiality, the timing of the follow-up and the types of questions to be asked.

Clients selected for evaluation should be asked to sign a written consent form that explains the purpose and methods of the follow-up procedures. A sample form is shown in Workbook 1, Appendix 2. The consent form should indicate the reason you are evaluating clients, the (random) process of selection, assurances of confidentiality, the timing of the follow-up and the types of questions to be asked. It also should indicate that the client has the right to decline to participate and that their decision will not influence current or future participation in treatment. The form records the client's name, address and telephone number and asks for details of other people who may be contacted to assist in locating the client. It is important to know if follow-up workers can, if necessary, identify themselves to others who may respond to the follow-up contact. In outcome evaluations, it is common practice to ask all clients to complete the consent form at intake and then take a random sample of those who agree.

For additional information and advice about preparing and using consent form, review Workbook 2 Step 1A, entitled "Manage Ethical Concerns."

Your consent form should accommodate the special circumstances of young clients whose right to consent to treatment and evaluation may need to be endorsed by parents or guardians. The legal requirement to obtain consent from parents or guardians will vary across jurisdictions: Check with your local authorities and/or an ethics board to determine the best way to proceed in your setting.

The process of locating former clients for follow-up can be time consuming and frustrating for follow-up workers. This is especially the case for socially unstable clients and those who may have relapsed. Interviewers must be tolerant and flexible. Whether face-to-face or telephone interviews are planned, a pre-determined schedule of contact attempts must be followed (e.g., five telephone calls at varying times of day).

5. The conduct of follow-up interviews

Many follow-up studies of clients of PSU services use telephone interviews. There is a general consensus among evaluators that telephone interviews can provide valid outcome data when properly conducted (IOM, 1990). They are a good option for collecting follow-up in settings where most clients have telephones. However, they are inappropriate in situations where few clients have phones, or where phone calls to clients' homes may violate their rights to privacy.

If telephone interviews cannot be used, your next best option is to interview clients on the premises. Arrangements other than this pose logistical difficulties, for example, finding suitable places to conduct interviews without compromising client confidentiality or posing risks to interviewers. Interviewers should not go to clients' homes or other street addresses except in pairs or with clear backup support. Otherwise, they may place themselves at risk. These are not trivial concerns and the safety of follow-up workers engaged in face-to-face follow-up should be given careful consideration.

6. Selection and training of interviewers

All data collectors should be trained thoroughly before starting work with clients.

An important standard for outcome evaluation is that all client interviews and data collection be undertaken by people not associated with the provision of the intervention. This is the case for both pre-treatment and follow-up interviews, although practical and resource constraints may make it difficult during the pre-treatment assessment. Resource constraints may require that programme staff assist in the collection of follow-up information. However, they should not do so for clients they have treated. This is important in order to avoid clients "faking good" at follow-up to the clinical staff who have treated them.

All data collectors should be trained thoroughly before starting work with clients. Review Workbook 2 Step 1C, entitled "Develop a Data Management Plan," for more information about how to do this.

If your follow-up interviews are to be conducted by telephone, the interviewers must have a professional manner and clear voices over the phone. If face-to-face interviews are used, interviewers should be selected and trained such that clients feel they can talk freely. For example, if you are planning face-to-face interviews and the people to be interviewed are young adults with unconven-

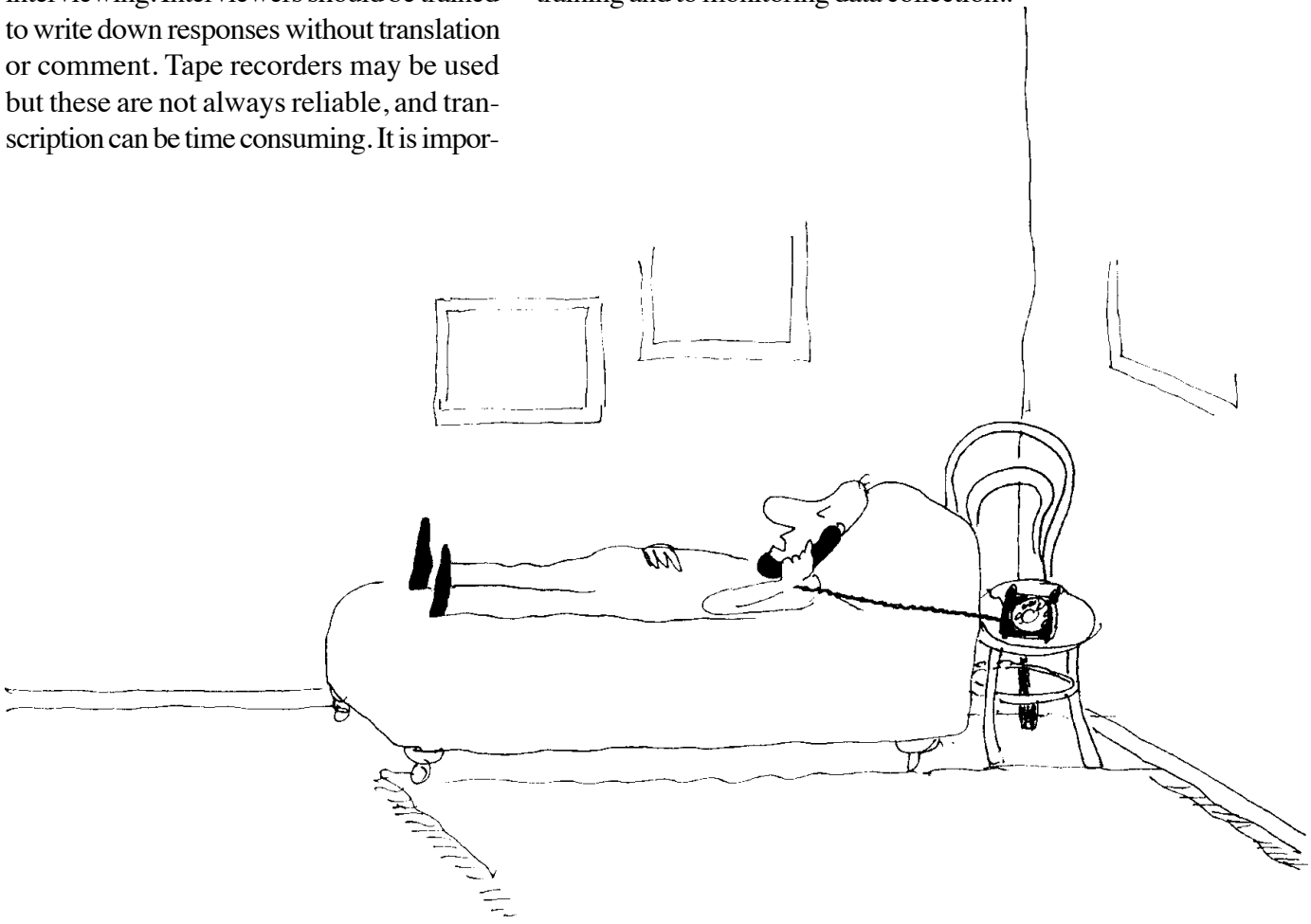
tional lifestyles, try to engage young interviewers who have flexible time schedules and who feel comfortable in casual clothes and conversation. Similarly, a face-to-face follow-up of adults or elderly persons would best be done by older, more conventional individuals. The gender of the interviewer may also be important, especially if your programme has objectives specific to female or male issues.

The language in which the follow-up interviews is conducted is of obvious concern. This may be difficult to accommodate in all cases, especially if your programme has a multicultural clientele. Careful attention must be given to the use of outcome measures validated in one culture and developed in particular language, and then translated into another language. Such cross-cultural application may significantly influence the reliability and validity of the measure.

Interviewers should be familiar with the interview schedule and objectives of the evaluation. They should practice before they start interviewing. Interviewers should be trained to write down responses without translation or comment. Tape recorders may be used but these are not always reliable, and transcription can be time consuming. It is important that interviewers be well supervised to ensure they stay within the agreed evaluation protocol and act professionally.

Whether you are using face-to-face or telephone interviews, an important issue in the selection and training of interviewers is the extent to which they are allowed to address clinical issues that may arise. It is recommended that a written protocol be developed for interviewers to guide their response to requests for additional treatment or more serious emergencies such as expressed suicidal ideation. While clinical training and experience are usually not required of follow-up workers, they must be capable of responding professionally and ethically to a range of situations that may present themselves.

Some programmes have trained former clients or other volunteers to locate clients and conduct the follow-up interviews. This may be an option for programmes with limited resources for outcome evaluation. In these cases, particular attention should be given to training and to monitoring data collection..



Choosing your outcome measures

Outcome measures can be selected from three broad domains:

- reduction of PSU
- improvement in personal and social function
- reduction in public health and safety risks

A wide range of potential outcome measures in each of these domains may be relevant for the evaluation of your treatment service or system. Your choice of outcome measures is critical to the success of your evaluation. The decisions you make are closely tied to decisions you will have to make regarding data collection procedures. For example, some measures will be appropriate for self-completion, others by telephone and still others may require a face-to-face interview with particular groups of clients. Most importantly, your choice of measures must be guided by the objectives that your treatment service or system is trying to achieve.

The table below identifies many possible outcome measures within each of these areas:

Domain: PSU

- Workbook 1, Appendix 2 includes a brief format for measuring quantity/frequency of PSU
- Timeline Follow-back Method (Sobell and Sobell, 1992)
- Alcohol and Drug Use Subscales of the Addiction Severity Index (McLellan et al., 1988)
- Quantity/Frequency Measures from Directory of Outcome Measures (Addiction Research Foundation)

Domain: personal and social functioning

- Short Alcohol Dependence Data (Raistrick et al., 1983)
- Drug Abuse Screening Test (Skinner, 1982)

- WHO - CIDI (Witchen, 1994)
- Symptom Checklist-90 (Derogatis, 1977)
- Beck Depression Inventory (Beck et al., 1961)
- Perceived Social Support (Procidano and Heller, 1983)
- Social/Family Subscale of the ASI (McLellan et al., 1988)

Domain: Public Health and safety risks

- Workbook 1, Appendix 2 includes a brief format for measuring HIV-risk behaviours, and health, social, and correctional services
- Legal Sub-scale of the ASI (McLellan et al., 1988)

These measures are presented as examples only. You must decide on their appropriateness and availability for your clients and your culture. In making your selection of outcome measures, you should consider:

- the objectives of your treatment service
- the client population you serve
- the time you are prepared to invest in your assessment process for the collection of data
- the potential use of a computer to assist in collection of the information (i.e., self-administered questions)
- the time period over which you wish to have clients report PSU
- established reliability and validity data for your culture
- cost to use the instrument if not in public domain
- the follow-up data collection strategy (i.e., telephone versus face-to-face interviews)
- the resources you have available for data collection, analysis and preparation of reports

In addition to your **outcome** measures, you also need measures that help you explain or **predict** outcome for certain groups of clients. For example, PSU is an outcome measure; while the number of days in treatment is a predictor variable.[1] Information on any given client?s participation in treatment needs to be linked with his/her outcomes. Demographic characteristics of clients such as gender, age, and socio-economic status are often used as predictor variables. Other predictor variables could include:

- severity of dependence
- extent of family and social supports
- psychiatric symptoms, in particular the presence of anti-social personality diagnosis

These predictor variables are measured by some of the questionnaires listed in the table above.

[1]Scientific jargon refers to these as dependent variables (your outcome) and independent variables (your predictors).

What information source should you use?

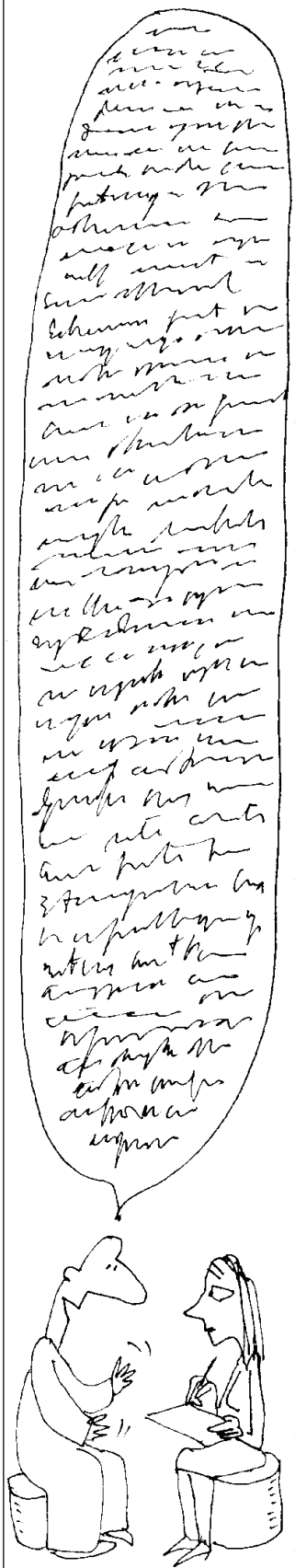
...ethical considerations prevent collecting data from third parties (e.g., family members) without clear, written permission from clients themselves.

You will have to decide whether to collect all of your outcome information from one source (usually the client), or from more than one source. Having additional information to back-up clients' self-reports is recommended (McLellan et al., 1996) (e.g., breathalyser; urine screening tests, and/or collateral reports). This may not be practical in all situations or if limited resources are available for evaluation.

Do PS users tell the truth? Overall, research indicates that self-reports of PSU, criminal and other behaviours *are reasonably reli-*

able and valid under certain conditions (Sobell et al., 1992). PS users are more likely to give accurate answers when:

- they are sober and PS free at the time of the interview
- confidentiality can be assured and there are no consequences for reporting PSU or illegal behaviours
- the interviewer is skilled and non-judgemental and there is good rapport between the interviewer and the respondent
- questions are clear, direct and easily understood by the respondent



There is additional evidence that PS users are more accurate when they are aware that their answers will be verified against third party reports or official records. Such verification is not always possible, but if family members can be interviewed or police records checked, clients will have less reason to deny their actions to you.

Not every client will tell "the whole truth and nothing but the truth" about every aspect of their lives in treatment outcome studies. Some will be motivated to under- or over-report certain behaviours and many will not remember everything they did, or everything that happened to them. However, if the conditions for the interview are right, few respondents are likely to present distorted accounts of their lives and behaviours over a given follow-up period.

Face-to-face interviews present opportunities to observe former clients and these observations may give clues about their lives and situations. Interviewers can be trained to rate clients with respect to their levels of intoxication, appearance and mood and signs of PSU, for example, smell of alcohol, presence of bottles or syringes and other paraphernalia, or needle marks on the arms.

Family members and friends may know about the lives and behaviours of former clients and be willing to report these to evaluators under certain conditions. However, ethical considerations prevent collecting data from third parties(e.g., family members) without clear, written permission from clients themselves. When this permission has been obtained, third parties approached for an interview should be told why the information is needed and what the consequences will be for the client. Reports from third parties will be most reliable under the same conditions as those identified for client interviews.

A serious limitation to third party interviews is that respondents are often unaware of the behaviours of individual clients. PSU may take place in private or away from family members and friends. Clients may not always tell others what they have been doing. Families and friends may have observed clients in various states of intoxication, however, and be willing to report this to a follow-up worker.

Records kept by police, hospitals, employers, welfare workers and other agencies may indicate contacts with former clients. These records may be accessible if clients have given written permission. The value of these records for treatment outcome studies varies with the type of clients treated. For clients who typically have contacts with criminal justice, health or social agencies records kept by these agencies may show significant changes following treatment. If clients typically have few contacts with these agencies, searches of records may turn up little value. The value of records for outcome measurement depends on their completeness and accessibility. When records are not computerised, or stored alphabetically, a great deal of effort may be needed to abstract relevant information on individual cases.

Recent alcohol use can be detected with a breathalyser or through urine and blood tests. Less recent use of certain psychoactive substances can be detected in urine, blood, and hair samples. Tests of these samples require supplies and equipment for collection of specimens. Access is also needed to properly equipped laboratories. The costs involved may be prohibitive for many programme evaluations.

Outcome evaluation at the system-Level

Recently, more attention has been given to assessing outcomes associated with large networks of treatment programmes. System-level outcome evaluation can involve any of the evaluation designs described above — randomised controlled trial, comparison group, or pre-post test. In practical terms, however, it is difficult to randomly assign clients to one network of services versus another. The most practical design to implement at the system-level is the pre-post design with the same data collected from all programmes in the defined network. Large scale outcome monitoring systems are now operational in the USA and Canada (e.g, Harrison et al., 1996; Policy and Service Consultation Information and Funded Services, 1995). Others are being developed elsewhere. The second case example located in the back of this workbook (by Gossop and colleagues) is a good example of an outcome evaluation at the national system level.

Practical issues are considerably magnified with outcome evaluations at the system-level:

- involvement of a wider range of key groups in the evaluation process and more difficulty achieving consensus on outcomes to be measured
- more involvement of funders and/or payers in developing the evaluation questions. They may, for example, find it difficult to formulate specific policy questions that they would like addressed
- difficulty identifying relevant outcomes across programmes with widely varying objectives and client populations (e.g., detoxification centres, assessment and referral centres, treatment programmes aftercare programmes, youth programmes, and programmes serving the elderly, the homeless, or multicultural populations)
- more difficulty getting system-wide buy-in to the evaluation process due to fear that the results will be used to restructure the system in dramatic ways and cut programs
- limitations of the pre-post evaluation design in attributing causality to outcomes obtained and fearing that the results will be used inappropriately for policy decisions
- fear among service providers that if outcomes are being measured in a **sample** of agencies the results may not be representative of their program

For these reasons, it is recommended that you consult with an experienced outcome evaluation researcher before attempting a project at the system level.

It's your turn

Put the information from this workbook to use in your own setting or treatment system. Complete these exercises below.

Remember to use the information from Workbooks 1 and 2 to help you complete an evaluation plan. Review that information now, if you have not already done so.

Exercise 1

Think about your treatment programme. List five **general areas** in which you want to know the effectiveness of your programme.

Example: Is our women's programme effective?

1 _____

_____

_____

2 _____

_____

_____

3 _____

_____

_____

4 _____

_____

_____

5 _____

_____

_____

Exercise 2

For each area that you listed above, choose a series of specific questions to ask:.

Example (from above):

A) Does our women's programme reduce PSU?
B) Does our women's programme reduce the severity of depressive symptoms?
C) Among participants, do younger or older women do better?

Now it's your turn. Follow the same procedure for each of the five areas that you listed in Exercise 1.

Exercise 3

Review the questions that you created in Exercise 2. Consider which of these questions are **feasible** to study, and which are **most important** to study in your setting. You should review Workbook 1, Evaluation Planning, for additional information about how to do this. Once you have settled on key questions, decide how you will measure each of them.

Review potential outcome measures using the appendix in this workbook, other resource manuals, and if possible, consultation with evaluators in the PSU field. Then:

A. Decide which of these data collection methods you will use:

	Yes	No
open-ended items on self-administered questionnaire		
one-on-one interviews		
focus groups		
program documents		
clinical observations		

B) Decide **how** data will be collected and **by whom**:

Baseline data	Follow-up worker	Follow-up method
blended with clinical assessment	independent evaluator	telephone
separate from clinical assessment	programme staff	face-to-face
but collected by programme staff	volunteer	mail out
collected by external evaluator	flexible depends on client	flexible depends on client

Example (from above):

Data will be collected using self-administered questionnaires and corroborating medical chart data:

a) PSU:

- Questions about frequency, quantity, and type of PSU over the past eight weeks
- Review of medical records for results of intake and discharge toxicology screens

b) Depressive symptoms: The Beck Depression Inventory

c) Age: General demographic questions

All data will be collected by an external evaluator, using face-to-face contact within the clinic building, and separate chart review.

Now it's your turn. Follow the same procedure for each of the questions that you listed in Exercise 2.

Exercise 4

Using the information provided in this workbook about how to design and conduct an outcome evaluation, make the following decisions:

- Choose an evaluation design:
  - experimental
  - comparison group
  - pre-post
- Choose a sampling procedure for choosing clients to survey
- Decide the timing of the evaluation
- Develop a procedure for ensuring clients' confidentiality and promoting their honesty in answering questions.

Example (from above):

- *Given programme resource limitations, a pre-post design will be used.*
- *All clients checking in for their assessment appointment will be asked by staff to meet with an external evaluator while waiting for their appointments. Pre-treatment data will be collected over a three month period of time.*
- *After agreeing to participate, clients will be given the questionnaires, and envelopes in which to place their completed questionnaires before returning them to the evaluator. Clients will be instructed to complete the questionnaires before leaving the clinic that day. The following statement will appear at the top of the questionnaire:*

  "Please help us improve our programme by answering some questions about your PSU and related problems. To ensure your confidentiality, please do not write your name on this form. When you are finished, place the form in the envelope (provided) and seal it closed, then give it to the evaluator in the waiting area."
- *ID numbers will be used in place of names on all questionnaires. A confidential list will be kept that links these ID numbers with clients? names and contact information. This list will be kept separate from the data, to further ensure client privacy. Three months after discharge from the 8-week programme, the clients will be contacted by the evaluator to schedule a follow-up meeting that is roughly 6 months post-admission. Clients will return to the clinic for this confidential follow-up meeting, where they will complete the same questionnaires and return them to the evaluator.*

Now it's your turn. Follow the same procedure for your evaluation questions.

Exercise 5

You will need to prepare a consent form that explains the purpose of your study. Review Section 1A of Workbook 2, entitled, "Manage Ethical Issues," for more information about the important topic of participants' rights in evaluation research. Also review Workbook 1, Appendix 2 for an example of an outcome evaluation consent form.

In general, all participants should be asked permission ahead of time before being enrolled in the study. When you do this, your should explain the purpose, nature, and time involved in their participation. No person should be forced or coerced to participate in the study.

The consent form should:

- describe the purpose and methods of the study
- explain what they will need to do if they participate
- explain that participation is voluntary

Now it's your turn. Using the example in Appendix A, and the information provided in Workbook 2, section 1A, write your own consent form.

Exercise 6

Run a pilot test of your evaluation measurement and procedures (including your follow-up interviews) to ensure that everything runs smoothly. Review section 1C of Workbook 2 entitled "Conduct a Pilot Test" for specific information about how to do this. In general, pilot tests assess these questions:

- Do the questions provide useful information?
- Can the questions be administered properly? For example, is it too long or too complicated to be filled out properly?
- Can the information be easily managed by people responsible for compiling the data?
- Does other information need to be collected?

Example (from above):

A pilot test will be run during one clinic day: 3 November. During this day, all patients checking in for an assessment appointment will be asked to complete the questionnaire. Afterwards, their responses will be examined to determine whether they seemed to understand the questions and were answering honestly. All persons involved with distributing the forms and tallying the data will be interviewed to determine their views on any improvements that could be made in the process and/or to the forms. Follow-up contact procedures and interviews also will be pilot tested.

Now it's your turn. Write down how you will pilot test your evaluation study. Don't forget to review Workbook 2 first!

Conclusion and a practical recommendation

In this workbook, we have outlined the basic principles and practices of outcome evaluation of PSU services and systems. The goals of this type of evaluation are the assessment of change within different dimensions of the client's life and demonstrating that your programme had a role to play in causing these changes. You have learned about the design of outcome evaluations — experimental, comparison group, pre-post — and how your choice of design affects the confidence you have in attributing changes in client's to their participation in your program. You also have learned about other methodological issues important for all types of outcome evaluation.

A word of advice: trade-offs always have to be made to the **rigour** with which you collect and analyse information to answer evaluation questions, and the **amount of resources** available to you. This is especially true for outcome evaluations. Your research goal should be to achieve the best possible information with the expertise and resources available in your setting. Be sure to review your resources carefully before embarking on an outcome evaluation.

After completing your outcome evaluation, you want to ensure that your results are put to practical use. One way is to report your results in written form (described in Workbook 2, Step 4). It is equally important, however, to explore what the results mean for your programme. Do changes need to happen? If so, what is the best way to accomplish this?

Return to the expected user(s) of the research with specific recommendations based on your results. List your recommendations, link them logically to your results, and suggest a period for implementation of changes.

Remember, outcome evaluations provide important information on the effectiveness of your programme. It is important to use the information to improve treatment services. Through careful examination of your results, you can develop helpful recommendations for your programme. In this way, you can take important steps to create a "healthy culture for evaluation" within your organisation.

References

Addiction Research Foundation. *Directory of Outcome Measures for Addiction Treatment Programs*. Available from Marketing Department, Addiction Research Foundation, 33 Russell Street, Toronto, Ontario, Canada, M5S 2S1, Undated.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. An inventory for measuring depression. *Archives of General Psychiatry*, 1961, 4:561-571.

Derogatis, L.R. *SCL-90: Administration, scoring and procedures manual-I for the revised version*. Baltimore: Clinical Psychometrics Research, 1977.

Harrison, P.A., Beebe, T.J., Fulkerson, J.A., Torgerud, C.R., The development of patient profiles for Minnesota's treatment outcomes monitoring system. *Addiction*, 1996, 91(5): 687-701.

Institute of Medicine (IOM), Committee for the Study of Treatment and Rehabilitation Services for Alcoholism and Alcohol Abuse. *Broadening the Base of Treatment for Alcohol Problems*, Washington, DC: National Academy Press, 1990.

McLellan, A.T., Grissom, G., Alterman, A.I., Brill, P., & O'Brien, C.P. (1993). Substance abuse treatment in the private setting: Are some programs more effective than others? *Journal of Substance Abuse Treatment*, 1993, 10:243-254.

McLellan, A.T., Laborsky, L., Caccida, J., et al. *Guide to the Addiction Severity Index:Background, administration and field testing results*. U.S. Department of Health and Human Services. Rockville, Maryland, 1988.

McLellan, A.T., Woody, G.E., & Metzger, D. et al. Evaluating the effectiveness of addiction treatments: Reasonable expectations, appropriate comparison. *The Millbank Quarterly,* 1996, 74(1):51-85.

Policy and Service Monitoring Consultation Information and Funded Services. *Outcome Monitoring System: Outpatient Treatment t Procedure Manual*. Edmonton, Alberta: Alberta Alcohol and Drug Abuse Commission, 1995.

Procidano, M.E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal of Community Psychology*, 1983, 11(1):1-24.

Raistrick, D., Dunbar, G., & Davidson, R. Development of a questionnaire to measure alcohol dependence. *British Journal of Addiction*, 1983, 78:89-95.

Skinner, H.A. The Drug Abuse Screening Test. *Addictive Behaviors*, 1982, 7:363-371.

Sobell, L.C., & Sobell, M.B. Time-line followback: A technique for assessing self-reported alcohol consumption. In: R. Litten & J. Allen (Eds.). *Measuring alcohol consumption: Psychosocial and biological methods*. New Jersey: Humana Press, 1992:41-72.

Sobell, L.C., Toneatto, T., Sobell, M.B., & Shillingford, J. Alcohol problems: Diagnostic Interviewing. In: M. Hersen & S.M. Turner (Eds.). *Diagnostic interviewing (*2nd ed.). New York: Plenum Press, 1992.

Witchen, H.U. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Psychiatric Research,* 1994, 28: 57-84.