



A brief guide to evaluation for NSW drug summit programs

A brief guide to evaluation

for NSW drug summit programs

March 2001

The Cabinet Office
(NSW Office of Drug Policy)

This guide was written by Lisa Zipparo, Evaluation Manager, NSW Office of Drug Policy, in consultation with the NSW Senior Officers Coordinating Committee on Drugs and the NSW Expert Advisory Group on Drugs.

Published by The Cabinet Office
New South Wales Government
©March 2001
I S B N 0 7313 1587

Table of Contents

Who should read this guide?.....	1
Why evaluate your program?	2
What do you need to do first?.....	3
What should you measure?.....	4
How do you identify short-, medium- and long-term outcomes?.....	6
Are long-term outcomes useful measures of program impact?.....	8
Are there solutions to the limitations of measuring long-term outcomes?	9
An example	10
What are the challenges in identifying short- and medium-term outcomes?.....	11
What tools can you use to measure outcomes?.....	12
How should you design an evaluation?.....	14
What skills are required to plan and conduct an evaluation?.....	16
How much should you spend?	17
What about ethical considerations?	18
Do you need assistance?	19
References	20



DRUG SUMMIT PLAN OF ACTION STATEMENT OF OBJECTIVES

THE NSW DRUG SUMMIT PLAN OF ACTION recognises the shared desire of people in New South Wales that we should live in a society free from drug problems.

The Plan has the following objectives:

1. To prevent the uptake of and addiction to illicit drugs, particularly by young people
2. To work in partnership with the community to understand and educate with the aim of reducing illicit drug use
3. To increase access to a comprehensive, high quality and innovative range of treatment and counselling services
4. To ensure that people who traffick in drugs are detected and penalised
5. To promote trials of treatment and support options, evidence-based practice and evaluation of all programs
6. To assist people into treatment and away from the criminal justice system and break the cycle of crime and drug abuse
7. To reduce the harm caused by risk taking behaviour associated with drug use
8. To reduce the impact on the community of drug related crime and anti-social behaviour
9. To reduce the supply of illicit drugs by promoting best practice intelligence-led policing and close cooperation with law enforcement agencies in other jurisdictions
10. To better equip health and welfare frontline professionals in providing care and management of people with drug problems

Who should read this guide?

- This guide is intended for those who are designing an evaluation, taking part in an evaluation, or preparing an evaluation brief for consultants to evaluate a NSW Drug Summit Program.
- This guide provides some key points to consider when embarking on a program evaluation.
- This is not intended as a definitive guide to evaluation. It aims to provoke ideas and raise some critical issues that are important to consider if a useful evaluation is to be planned and carried out.
- There are many detailed guides to evaluation. Many are available on the internet. Some useful guides have been referenced at the end of this guide.

Why evaluate your program?

A well-designed evaluation will help you to:

- set clear aims and objectives for the program
- ascertain the progress of the program
- develop and improve the program
- guard against the program becoming stagnant
- test theories about why the program should work
- identify problems and generate solutions during the life of the program
- develop a scientific basis for decision-making
- gather credible evidence about the program's effects
- assess if the program has the intended impact
- set future directions.

What do you need to do first?

Together with key stakeholders in the program:

- Establish a clear description of the program that is agreed upon by key stakeholders. Start by clearly identifying:
 - program goals
 - strategies to achieve program goals
 - populations being targeted by the program
 - expected effects of the program
 - any possible unintended effects of the program
 - any possible external factors that may affect the program.
- Develop a step by step description of what the program aims to accomplish and the activities that will be employed to accomplish those aims.

Design an evaluation plan that:

- Clearly states the aims of the evaluation and how the aims will be met.
- Identifies the key questions you want the evaluation to answer.
- Identifies key questions that others have about the program effects.

What should you measure?

Distinguish between process measures, output measures and outcome measures:

- **Process measures** are anything that is done, offered, or created by the service provider. These are not outcome measures.
- Process measures might include descriptions of activities, treatments, steps taken to implement a program, partnerships or collaborative relationships formed as part of the program implementation.
- **Output measures** seek to quantify the amount of work undertaken. They are usually measures of quantity, for example: the number of people who accessed the program and the number who completed the program.
- Outputs are useful in defining what a program produces.
- Outputs are the vehicles for producing the program outcomes.
- Output measures alone do not indicate whether the program goals have been achieved, nor do they reveal much about the quality of the program.
- **Outcome measures** seek to assess whether the program has achieved its intended goals.
- Program outcomes can be defined as the intended changes in an individual, which are attributable to the program's intervention¹. The intended changes can involve health status, knowledge, attitudes, behaviour (for example).
- All programs have long-term as well as shorter-term outcome goals which lead to the long-term outcomes.

¹ This definition was adapted from a definition of health outcomes coined by Teeson, and associates (2000). The reader is referred to this publication entitled 'The Measurement of Outcomes in Alcohol and Other Drug Treatments' for a comprehensive review of health outcomes measurement.

- In order to be able to evaluate the impact of a program, the short-, medium- and long-term outcome goals of the program must be identified.
- It is important to identify the short- and medium-term outcome goals of a program since it is sometimes difficult to assess the impact of a program upon longer-term goals.

It is also important to describe **program inputs**. They include:

- anything that is required to carry out the program including personnel, time, money, equipment and any other resources required.
- It is important to describe program inputs in order to be able to identify any mismatch between available and required resources and to enable economic evaluation.

How do you identify short-, medium- and long-term outcomes?

Together with key stakeholders in the program:

- Build a map of the program that illustrates, step-by-step, the program activities and the short-, medium- and long-term outcomes each activity is expected to produce.
- It might be easier to start at the top by stating the long-term goal of the program and working down to the program activities and the short-term goals they aim to achieve. Long-term outcomes are often the easiest to identify as they are usually the reason why the program exists.
- Examples of long-term outcomes for drug prevention programs include:
 - decrease in use of illicit drugs
 - decrease in number of fatal overdoses
 - decreased incidence of injecting drug user related cases of HIV, HCV
 - decrease in drug-related crime.
- Once you have developed a step by step description of what the program aims to accomplish and how, you will be able to more easily identify the short-, medium- and long-term goals of the program in order to develop short-, medium- and long-term outcome measures.
- The diagram overleaf provides an example of a map that identifies short-, medium-, and long-term outcomes for a hypothetical needle exchange program. The diagram is a very simplified version of what an actual program map might look like. The diagram does not describe the tools that might be used to measure each of the outcomes. Potential tools are discussed later in this guide.

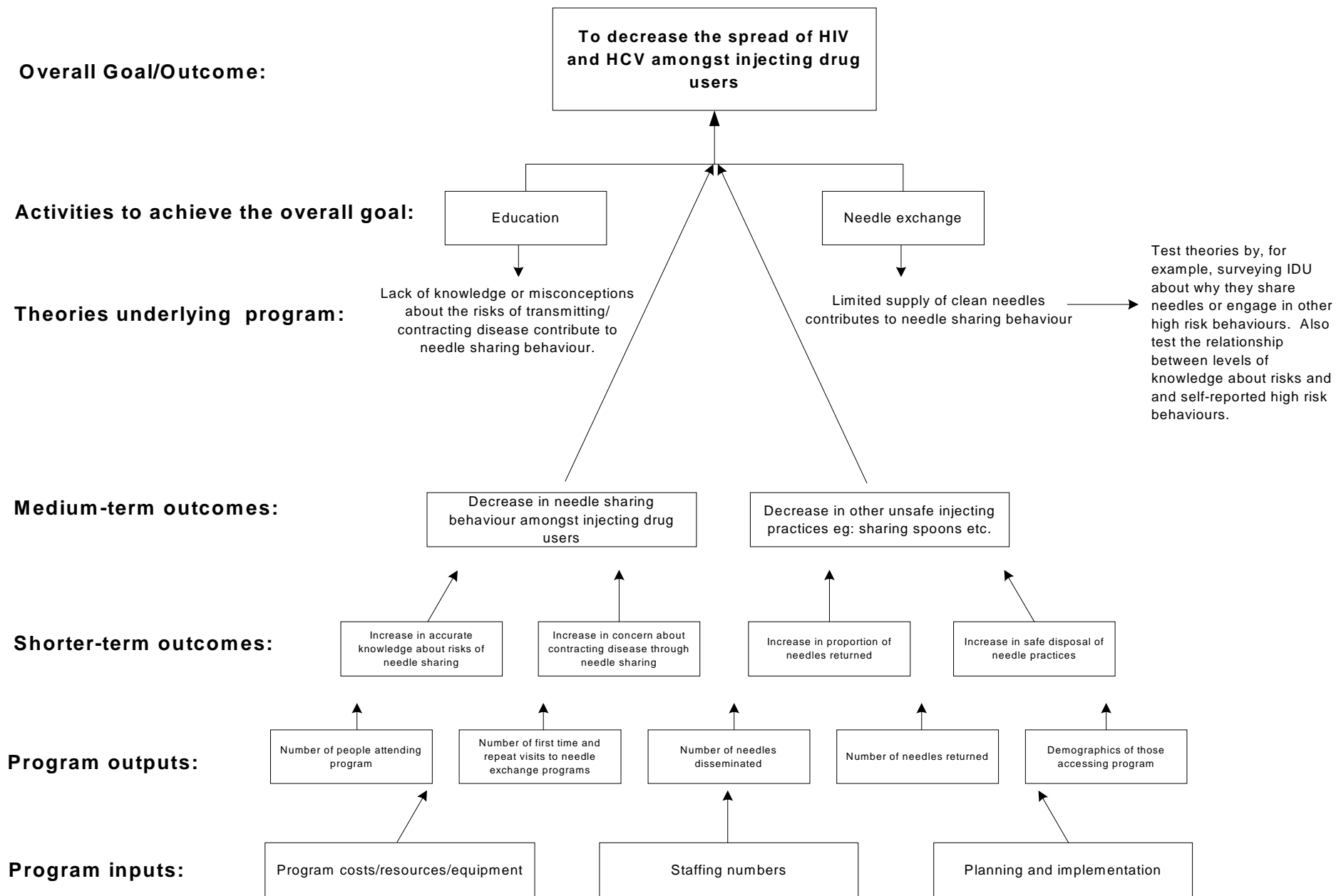


Diagram 1: Model for a step-by-step description of outputs and outcomes for a hypothetical needle exchange program

Are long-term outcomes useful measures of program impact?

- It can be difficult to assess the impact that a program may have on long-term outcome measures for the following reasons:
 - the potential long-term nature of such indicators means that there may be a considerable time lag from the program intervention to the outcome measure actually showing a change;
 - it is often difficult to conclude that an individual program or even a group of programs have caused changes in overall (long-term) outcome measures because there are many potential external factors over and above intervention programs which may also impact on the outcome measure;
- Measuring long-term outcome alone can be of limited practical application to individual programs because they cannot necessarily tell us whether individual programs are working and why.

Are there solutions to the limitations of measuring long-term outcomes?

- While it is important to monitor long-term outcome measures as they are often useful indicators of the overall environment, it may be more useful for individual programs to also identify more immediate outcomes of the specific program being evaluated.
- Once you have built a step-by-step picture of your program's activities and the specific goals of each activity you are more easily able to identify short- and medium-term outcomes of your program.
- Together with key stakeholder of the program think about:
 - each of the most important activities of the program
 - how these activities are expected to bring about change
 - the theories behind why an activity is expected to work
 - any empirical evidence that supports those theories
 - the judgement of those who conduct the program as to how and why the activity can bring about change.
- Having a clear understanding of why a program activity is expected to bring about change, you can then construct measures that test whether in fact change is happening as predicted by the underlying theory.

An example

The needle exchange program provides an example of how this sort of evaluation might work:

The overall goal of needle exchange programs is to reduce the spread of blood borne diseases caused by needle sharing amongst injecting drug users (IDU). The theories as to why this program should achieve a drop in the spread of disease might include that needle-sharing behaviour is due to the lack of availability of clean needles and a lack of knowledge about the risks of disease transmission. If clean needles and education about the risks of needle sharing are made available, the assumption is that needle sharing behaviour should decrease resulting in a drop in the spread of disease.

Attempting to ascertain the success of the program by measuring the overall intended outcome of reducing disease transmission rates may not provide an accurate indication of the program's effects because disease rates may not be immediately sensitive to the program's effects. In addition, there are probably many other factors outside of the program's impact that effect disease rates.

Rather than putting all of your resources into measuring a long-term indicator, that may not be an exclusive indicator of the program effects, shorter-term outcomes that would predict the long-term outcome can be measured. These might include exploring reasons for needle sharing, measuring attitudes to needle sharing, knowledge about risks, needle sharing and other high risk practices, and testing for any changes pre- and post- program (see diagram on page 7).

What are the challenges in identifying short- and medium-term outcomes?

- Seeking to identify short- and medium-term outcome measures by thinking about the theory underlying the program assumes that plausible, attainable and testable theories of change underlie a program.
- This approach will encounter difficulties if the theoretical links between the short-, medium- and long-term outcomes are not well understood.
- Change can come about as a consequence of a number of factors. The challenge is to identify the most important factors that are likely to cause the desired change and to test those.
- While short- and medium-term outcome measures may not be as vulnerable to external factors as are long-term outcomes, shorter-term measures can still be impacted upon by factors other than the program effects. It is important to think about and identify possible external factors that may impact on any outcome measures you employ so that control measures or, at the very least, alternative explanations are available.
- Once you have identified what you want to measure, you have to consider how you are going to measure it.

What tools can you use to measure outcomes?

- There are a number of tools available to collect information about whether an outcome has been achieved. Information can be quantitative or qualitative.
- **Quantitative data collection tools** include surveys which can be:
 - self-reporting surveys that ask those directly involved in a program about their knowledge, experiences or attitudes; or
 - observer surveys that might seek the perceptions of family members or other relevant people other than the person involved in the program.
- Surveys can be designed to obtain qualitative as well as quantitative information.
- Surveys that collect quantitative information can be useful tools for measuring changes in knowledge or attitudes if they are administered prior to the commencement of a program and then after the program has finished. Differences in pre-program and post-program scores may provide useful information about program effects.
- Using the example of the shorter-term outcomes identified in the hypothetical needle exchange program map on page 7, pre- and post-education program surveys could be conducted to evaluate if the program has been successful in increasing knowledge about needle sharing risks.
- As well as designing a survey that is tailored to a program, there are a number of pre-existing methods of psychosocial assessment that may be useful tools for measuring mood states, personality characteristics, or self-perception (for example psychological tests).

- It can be challenging to find ways of measuring behaviour which do not rely on self-report methods. It may be possible to employ some form of clinical, laboratory, or other assessment that can provide an objective measure of a person's behaviour.
- Referring again to the example of the hypothetical needle exchange map on page 7, one of the medium-term outcome measures identified was a decrease in the rates of viruses found in blood taken from returned needles. Testing the blood in needles is an example of an objective method of trying to determine how widespread needle sharing behaviour may be.
- **Qualitative data collection tools** might include:
 - observation and case studies
 - focus groups
 - one-to-one interviews.
- Qualitative information is useful for gaining impressions about broad issues and experiences, but does not really allow you to quantify change.
- Gathering qualitative information can be a useful first step in obtaining ideas for questions you may want to ask when attempting to quantify change.

How should you design an evaluation?

The most critical question that an evaluation seeks to answer is usually what would have happened if the program did not exist.

- Identifying and measuring short- and medium-term program outcomes alone do not assist in being able to comment on what would have happened if the program did not exist.
- There are a number of ways that an evaluation can be designed to try and answer this question.
- The nature of the evaluation design will depend, in large part, upon the nature of the program being evaluated.

Some examples of evaluation designs include:

- Comparing measures for those who have attended the program (program group) with a group who have not attended the program (control group).
- Any differences between these two groups might indicate program effects.
- A significant problem with this type of comparison occurs when you are unable to randomly assign people to the program group and the control group. Because random assignment is rarely able to be achieved in program settings, it can often be the case that those who attend a program have pre-existing differences from those who do not take part in the program. Any differences in outcome measures detected between the two groups might, therefore, be a result of the pre-existing differences rather than because of the program effects. For example, an evaluation of needle exchange programs in the US found that the probability of HIV infection amongst those attending needle exchange programs was 33% compared with only 13% for those not attending needle exchange programs. This led some to the incorrect conclusion that needle exchange programs worsen the HIV problem.

Further analysis, however, showed that those who sought out needle exchange programs were more likely to engage in 'high risk' behaviours in the first place and were, therefore, at higher risk of contracting blood-borne diseases than those who did not seek out such programs². This difference would make comparisons between the two groups difficult to interpret in terms of program effects.

- An alternative to comparing a program group with a 'control' group is to utilise a repeated measures design in which measures are taken pre, during and post program intervention (for example) for each individual. The aim is to look for changes in an individual over time. When and how often measures are taken would depend upon factors such as the type and number of interventions and how often each participant is available for testing.
- How often measures are taken should be guided by a number of considerations:
 - keeping data collection to a minimum so as not to produce excessive data processing requirements;
 - guarding against creating a heavy schedule of testing for program participants;
 - the possibility that drop-out and non-compliance rates among program participants may make data collection challenging.
- The number of people who participate is critical to the success of your evaluation. Ideally, all those who participate in the program should have the opportunity to comment on the program and its effects. If this is not practical, then participants in the evaluation should be chosen at random (e.g. by pulling names out of a hat, or choosing every third person who comes in the door). If evaluation participants are not chosen at random, then you will not be able to generalise the findings to the entire program.

² Voelker, R. (1998) Do Needle Exchange Programs Work? The Robert Wood Johnson Foundation Quarterly Newsletter, Issue 3.

What skills are required to plan and conduct an evaluation?

Conducting useful evaluations requires the following:

- The ability to facilitate groups of stakeholders in order to design and document a program framework that describes the main activities and identifies short-, medium- and long-term outcomes of the program.
- Skills in designing data collection tools such as surveys.
- Skills in conducting interviews or focus groups.
- Knowledge of quasi-experimental design such as time series designs.
- Knowledge of databases to allow data input and analysis.
- Knowledge of statistical databases such as SPSS and SAS to allow statistical analysis of data might be useful if more in depth statistical analysis is required
- Even if you plan to employ a consultant to design and/or conduct the evaluation, it is important to be able to have access to staff with knowledge in the above areas to enable you to:
 - guide the consultant as to what you want from the evaluation
 - judge the quality of consultants' submissions and work.

If you need assistance in planning an evaluation and don't feel that you have access to people with the appropriate skills, page 19 has a list of contact people that may be able to assist with ideas and advice.

How much should you spend?

- The standard budget for evaluations is suggested at between 5% and 10% of the entire program budget.
- Some might say this is money better spent on program delivery and that the money spent on evaluation should be the bare minimum.
- Allocating funds to evaluation should not be regarded as a 'necessary evil' but rather an integral part of the program. There is no point in spending money on a program if you can't then adequately demonstrate that the program has had an effect.
- The key is to create an evaluation that not only reports on program effects but also gathers information that helps to shape and improve the program.
- Employing consultants to design and carry out evaluations can be a costly exercise. Costs can be kept down if consultants are given defined tasks. Rather than giving consultants a broad brief to "design and carry out an evaluation", ask for defined tasks, such as to "design a survey to test the pre-program and post-program attitudes of those accessing the program". Taking time to think about what you want from an evaluation will save time and money in the long run.

What about ethical considerations?

It is important to consider ethical issues when planning and carrying out an evaluation. Ethical issues include:

- The confidentiality and anonymity of all information obtained about and from participants. It is important to keep in mind that “the law cannot be relied upon to protect the confidentiality of information about research subjects or data if an application for a subpoena is made”(p.211)³.
- Obtaining informed consent from all participants.
- Obtaining permission for participants who are under 18 years of age.
- Consideration of any culturally sensitive issues.
- Appropriate and effective information provision for participants of non-English speaking background.
- The appropriate and confidential storage of all data.
- Your evaluation plan may require the approval of an ethics committee. You should check with the Research and Evaluation section in your overseeing agency about ethics approval requirements.
- For an in-depth discussion of ethical issues in research you can begin by consulting the *National Statement on Ethical Conduct in Research Involving Humans (1999)*, a National Health and Medical Research Council (NHMRC) publication which can be found at:
<http://www.health.gov.au/nhmrc/publicat/humans/contents.htm>.

³ Dixon, D. (1997) Ethics, Law and Criminological Research. The Australian and New Zealand Journal of Criminology, 30, p211-215.

Do you need assistance?

If you require advice about how to identify appropriate outcome measures for your NSW Drug Summit project, contact the person from your agency listed below, or contact the NSW Office of Drug Policy Evaluation Manager Lisa Zipparo on (02) 9228 4248.

NSW Department of Health

Devon Indig
Surveillance Officer
Drug Programs Bureau
02 9391 9220
DI NDI@doh.health.nsw.gov.au

NSW Police Service

Michael Cain
Drug Programs Coordination Unit
02 9339 5834
cain1mic@police.nsw.gov.au

References

Framework for program evaluation in public health. Morbidity and Mortality Weekly Report, Vol. 48, NoRR11;1;1-40. Koplan, J.P. (1999)

http://www.cdc.gov/mmwr/mmwr_wk.html

20/9/00

User-friendly handbook for mixed method evaluation (1997)

<http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/start.htm>

20/9/00

New approaches to evaluating community initiatives.

<http://www.aspenroundtable.org>

21/9/00

Evaluation strategies for human services programs. A guide for policymakers and providers. Harrell, A.

http://www.bja.evaluationwebsite.org/html/documents/evaluation_strategies.html

26/9/00

Evaluation treatments for alcohol and other drugs. Monograph Series No. 14 (1991). Fiona Mcdermott, Margaret Hamilton, Bruce Lagay.

<http://www.health.gov.au/pubhlth/publicat/document/ncada14.pdf> 27/9/00

BJA Evaluation Website

<http://www.bja.evaluationwebsite.org> 24/11/00

The measurement of outcome in alcohol and other drug treatment: A review of available instruments. NDARC Technical Report No 92. Teeson, M., Clement, N., Copeland, J., Conroy, A., & Reid, A. (2000)