Criminal Justice Evaluation Framework

Guidelines for evaluating criminal justice initiatives

Queensland Government

TABLE OF CONTENTS

INTRODUCTION TO THE CRIMINAL JUSTICE EVALUATION FRAMEWORK

These materials are provided to help Queensland criminal justice agencies evaluate their programs and initiatives by presenting a flexible framework for planning, implementing, and reporting on evaluations. The Criminal Justice Evaluation Framework (CJEF) outlines key evaluation questions, examines the type of data that needs to be collected to answer those questions, and provides information on how to manage the evaluation process.

The CJEF will be a useful tool for Queensland government officers who commission, prepare, and conduct criminal justice initiative evaluations, whether they are new to evaluation or already have evaluation experience. The guidelines are intended to complement relevant frameworks, policies, procedures and templates that individual agencies have developed for evaluation or project management.

This framework focuses on initiative evaluation. Initiative evaluations help program managers understand how an effort is working so they can make informed changes to elicit improved outcomes. The term initiative is deliberately broad, and is used to describe any set of programs, procedures, activities, resources, policies, plans, products, services, systems or strategies that aim to achieve common goals or objectives.[1]

Rather than taking a prescriptive approach to criminal justice program evaluations, the CJEF encourages the evaluation project team to use good judgment in matching the scope and methods of evaluation with the objectives of the criminal justice program and the requirements of specific evaluations.

Structure of the document

This introduction sets the scene, and looks at why evaluation is an important aspect of criminal justice programming and resource allocation.

The main part of the CJEF looks at the actual process of evaluation – the how to do it. These Guidelines for evaluating criminal justice initiatives provide a step-by-step approach to help the planning and implementation of efficient, theoretically and methodologically sound evaluations of criminal justice initiatives. This planning process is broken down into the following steps:

A glossary, list of resources and appendices are provided to guide evaluators through these steps.

What is evaluation?

Evaluation is the systematic collection and analysis of information to make judgments, usually about the effectiveness, efficiency and appropriateness of a program or initiative.[2] Although many types of evaluation exist, the process typically involves comparing aspects of your program and its impact to

expectations in order to judge the success of the program. Given that the majority of criminal justice initiatives are implemented in a non-controlled environment, however, a number of factors external to the initiative also need to be considered.

As is discussed further in the CJEF, the purpose, audience and significance of the initiative will influence the scope and type of an evaluation. For instance, a small scale evaluation conducted by those managing the initiative may be based on a few key evaluation questions and rely upon simple data collection methods such as participant feedback questionnaires. Alternatively, multiple methods may be utilised by external evaluators to conduct a large scale evaluation study based on a detailed evaluation framework and project plan.

Why evaluate?

Effective evaluation of criminal justice programs can provide considerable benefits, including:

- providing evidence of a program's level of achievement, or the **impact** that the program has on the targeted behaviour or population;
- developing an understanding of the **relationships** between the existing environment an initiative operates within, the initiative's actions or activities, and the objectives it aims to achieve;
- improving **planning** and **decision-making** by identifying the most effective aspects of the program and any barriers to success;
- demonstrating how effectively **resources** have been used;
- **attracting resources** for future programs (even if the evaluation shows that you haven't met your objectives, it shows that you are aware of the mistakes made and can identify more effective means of achieving outcomes);
- contributing to research and **best-practice** evidence, allowing learnings to be applied to other programs where appropriate; and
- promoting **accountability** for publicly funded initiatives.

Situating evaluation with program development

Used properly, evaluation is a dynamic process that assists in the ongoing management of programs. Ideally, evaluation should be designed at the time of project planning and form part of the on-going refining of program activities. However, in some cases this does not occur and evaluation is then undertaken as an afterthought at the end of a program. Unfortunately, because important data has not been collected, these evaluations may be unable to determine how well the program worked. By building evaluation into the early stages of program planning, it can examine the program throughout its life. This way, evaluation becomes part of the on-going management and refinement of criminal justice programs.

Figure 1 demonstrates how evaluation can be built into program development and implementation. The first step of program development is to conduct a needs assessment to identify the nature of the problem and the extent of need for the intervention. Data is gathered to determine gaps between the current state of affairs in a particular situation and the desired or optimal state. Program managers and

other stakeholders can then determine whether there is a gap and if so, what type of program should be developed to addressed these gaps. Needs assessments may also ascertain the characteristics of the target population for the initiative, and any assets in the local context that can be built upon.

Key measures of success must be determined when the program is being designed as well as provision made for the collection of relevant information to report against these measures. An ongoing program of monitoring key operational performance indicators should be put in place to ensure the program is tracking as anticipated. This may be as simple as monitoring the number of participants through a program or how resources are being allocated. Depending on the questions to be answered, the program can be evaluated or assessed at a suitable time after implementation. The evaluation results can then be used to revisit and reassess the needs and gaps identified initially, thereby forming a feedback loop of continuous improvement.

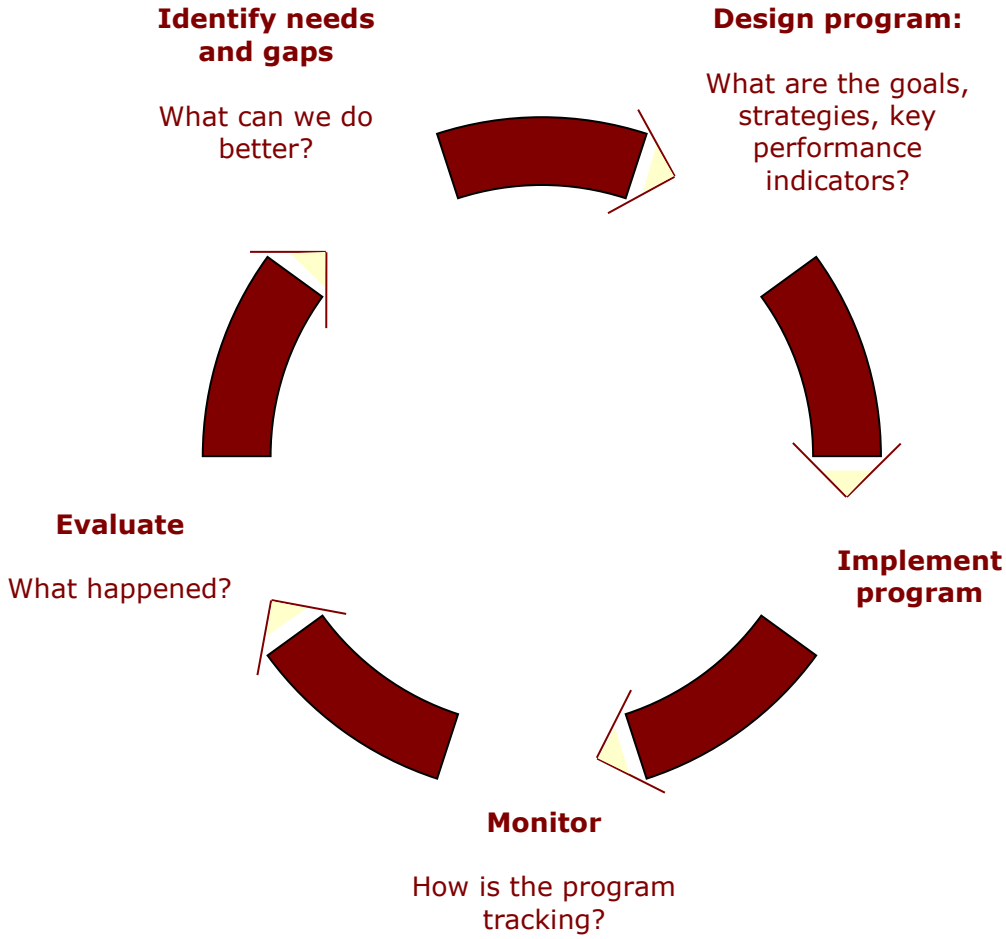

Figure 1. Evaluation: A process for ensuring continuous improvement

GUIDELINES FOR EVALUATING CRIMINAL JUSTICE INITIATIVES

Any evaluation must be planned and actively managed. You'll hear the word evaluation used in many contexts. Almost anything can be evaluated, and there are many different types of evaluation. Even though evaluation is used for different purposes, carried out at different times and at different levels, the basic process remains the same. The guidelines described in this document provide information that may be useful when planning your evaluation.

Figure 2 outlines a process that will help you to design a quality evaluation, and the remainder of this section provides more detailed information on this process. While the process, as presented diagrammatically, is sequential, in reality the process is more fluid and amenable to change. The evaluation model and process will vary depending on the purpose, scale and scope of the evaluation. The CJEF presents a flexible approach to evaluation planning, and the process should be tailored to suit your particular evaluation.

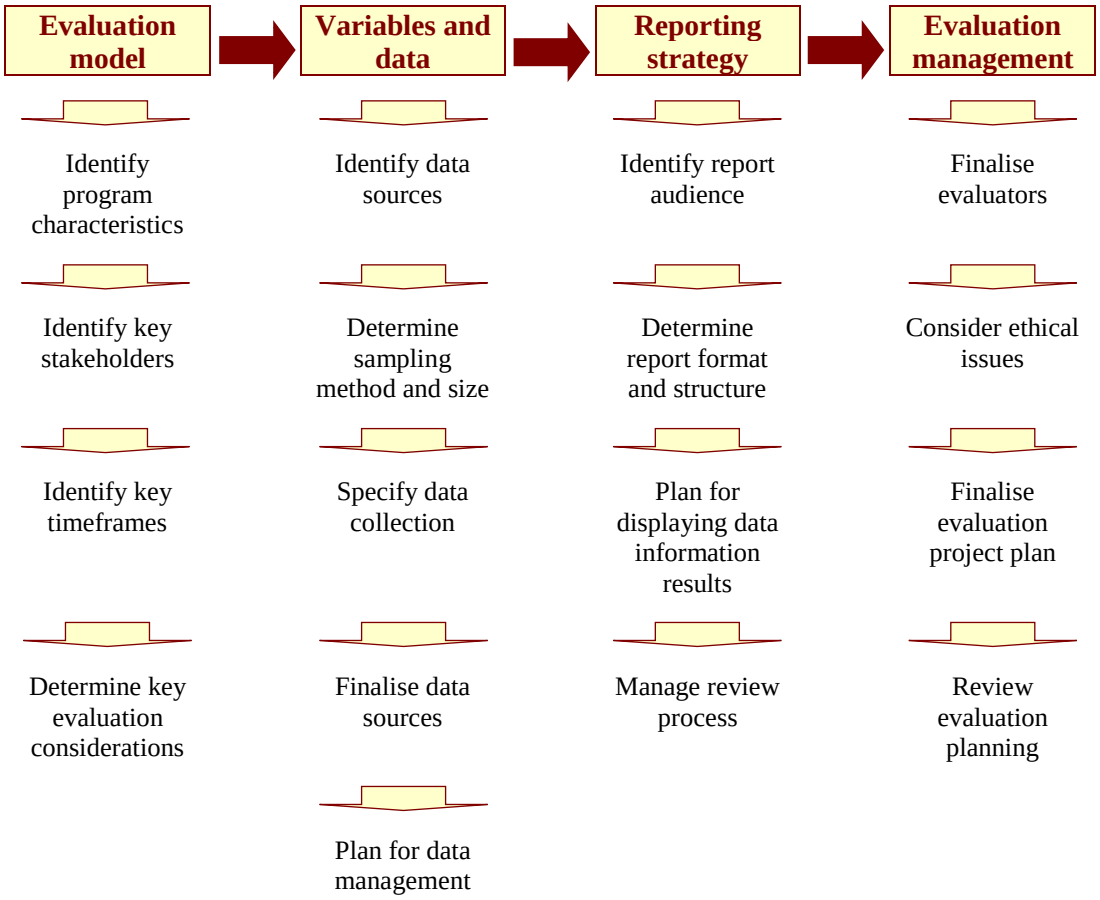

Figure 2. Process for planning an evaluation

Evaluation model

This section outlines the considerations that will inform the evaluation model. In evaluation, typically a combination of information sources is valuable for informing overall program assessments and estimates of program effects. These sources include intended program objectives; measures of performance; linkages between program inputs, outputs and outcomes; various environmental factors; the presence of external initiatives; the characteristics of the targeted population.

The most appropriate evaluation model is a function of various program characteristics and evaluation considerations, and informed by knowledge of key stakeholders and evaluation timeframes.[3] Quality evaluators will take full advantage of the wide range of alternative evaluation models that result from the interplay between these factors. When certain program characteristics are taken into account, the evaluation type, questions, methods and analyses that you select are likely to be appropriate and adequate. In turn, the evaluation is most likely to yield information that will be useful to decision-makers and other stakeholders within appropriate timeframes.

Identify program characteristics

A complete and detailed program description helps to focus the evaluation task. Only once an evaluation frame of reference has been developed can a choice concerning an appropriate research design be made. Identifying and documenting the theoretical approach, program type, and program specification provides a straightforward way of extrapolating the characteristics of criminal justice programs. Care must be taken to contextualise these characteristics within the social, cultural and political contexts that impact upon the program.

Theoretical approach

Evaluations need to take into account the mechanisms through which effects are assumed to be determined.[4] Essentially, this is the rationale which underlies the program design. Defining a theoretical approach involves specifying the underlying causal mechanisms behind the initiative.

For example, different theoretical approaches may have varying perspectives on the role of the individual environment, the community or the Government in preventing or reducing crime. In relation to crime prevention, four approaches predominate:[5]

- criminal justice approaches emphasise deterrence and incapacitation;
- situational approaches attempt to reduce the opportunities for offending by manipulating the immediate physical or social environment;
- community approaches focus on larger environments such as institutions and seek to minimise the social and organisational factors linked to crime; and
- developmental approaches emphasise intervening early in pathways that lead to antisocial or offending behaviour.

Program type

There are many different types of criminal justice programs, distinguishable by their main purpose and associated outcomes. For example, programs can be distinguished between those which seek to influence behaviour (for example public education programs, regulatory programs, case management programs) and programs that provide products or services (such as security services). Depending on the purpose of the program, each program will aim to achieve distinctive outcomes. Logic models (see below) are particularly useful for identifying the desired short to long-term program outcomes.

Evaluation strategies should be designed to correspond with the program outcomes that they wish to measure. Evaluation tasks for programs which seek to influence behaviour, for example, may include determining the type of people involved, and the extent to which people involved exhibit changes in action or behaviour. Thus, different types of programs, each with their distinguishing patterns of outcomes, necessitate distinct evaluations.

Program specifications

Program specification is usually expressed in terms of the:

- program **setting** (e.g., rural, urban, central business district);
- composition (e.g., Indigenous offenders, substance-abusing offenders, general public) and size of the **target group**; and
- type of need or **problem** being addressed (e.g., property crime, alcohol-related violence).

Again, an evaluation model which might be appropriate for a program that operates in a remote setting and has a significant proportion of Indigenous participants may be inappropriate for the same program operating in a larger centre with a small Indigenous population.

Developing a logic model may aid the conceptualisation of program characteristics and, in turn, designing the evaluation. Developing a logic model will clarify program activities and desired outcomes, and build consensus among program managers and stakeholders by connecting program activities with their intended short to long-term outcomes. The logic model format illustrated in Appendix A of the CJEF contains six core components: program goals and objectives, environment factors, and assumptions; and evaluation inputs, outputs, and outcomes.

Identify program goals and objectives

While evaluations are undertaken for a number of reasons, most attempt to measure the success or effectiveness of the program. Clearly, then, correctly identifying program goals and objectives is critical to successful evaluation because this information guides the development of evaluation questions and variables which will be used to measure the program's performance.

It is important not to confuse goals and objectives. A goal is a simple statement, which sets out the purpose of the program. Objectives are specific statements that are measurable and state exactly what you want to achieve – the desired outcome of a program. Objectives are a key tool for successful

program management and evaluation. You may need to clarify how goals are measured, and what is meant by terms such as needs, standards, effectiveness, and efficiency.

Where a project plan exists, determining the goals and objectives should be simple. If a plan has not been developed, or has not defined SMART objectives (see **Text Box 1**), it is possible, although quite complex, to define the objectives retrospectively for the purposes of planning an evaluation. This may involve talking to policy makers, managers and staff about the goals and objectives of the program. If you believe that the program, as it operates on the ground, differs from the original policy intent, then you should similarly seek to ascertain the "real" goals and objectives of the program. It may be important to contrast the policy intent or commonly understood objectives with the reality of the program.

It is important that evaluators set a realistic number of objectives to assess. Any more than three or four objectives can make a program difficult to evaluate. Programs that are delivered by more than one agency, or with whole-of-government outcomes, can similarly be difficult to evaluate. Sometimes, these programs can have many and varied objectives and, in some circumstances, objectives may even conflict. In this situation, it is very important to bring all relevant stakeholders together at the outset to determine which objectives should be evaluated (refer to Identify key stakeholders).

Text Box 1: SMART objectives

Objectives should be SMART, which means they are:

- ***S****pecific:* what will be achieved is clearly defined.
- ***M****easurable:* the outcome of an objective should be measurable.
- ***A****chievable*: the objective should describe something that can realistically be achieved within the timescale and resources set for the program.
- ***R****elevant:* the objective is essential to the broader aims of the program
- ***T****imebound*: a timescale has been identified for when the objective is to be achieved.

Identify program assumptions

Program assumptions are the beliefs we have about the program, the participants, and the way we expect the program to operate. They are the theories or underlying beliefs, validated with research and experience, on how the program will achieve success or the principles that guide your work. Assumptions underlie and influence the program decisions that are made. In developing a logic model, it is necessary to make explicit all implicit assumptions so that they may be explored and discussed. Often, inaccurate or overlooked assumptions are the basis for not achieving expected outcomes. Continue to check or clarify assumptions as evaluation planning progresses. Clarifying assumptions demands knowledge of the research or best practice in the substantive area, as well as common sense.

Identify evaluation inputs

Inputs are the resources, contributions and investments that are available for a program. The inputs used to produce the outputs may be financial, material or the amount of time that is committed. In program management it is important to be aware of exactly what resources are available to carry out the work. When resources are limited the objectives of a program and the scope of the work carried out can be affected. As well as measuring the success of a program, you could also be measuring the cost-effectiveness of any input and whether any specific methods or processes were particularly useful. You may find that a program or method was effective because of the level of resources available.

Identify evaluation outputs

Outputs are the activities, services, events or products of the program that reach those who participate or the targeted population. The outputs are not necessarily the final purpose of the program. Outputs are usually things that need to be done in order to produce the desired result. For example, anger management programs (output) may be delivered in order to achieve a decrease in violent behaviour (outcome). During the life of a program, the outputs should be monitored to make sure that they are being delivered on time and within the resources available.

Identify evaluation outcomes

The overall result of applying the inputs to a program and achieving the outputs is known as the outcome of a program. Examples of outcomes include changes in knowledge, skill development, changes in behaviour, capacities or decision-making, and policy development for individuals, groups, communities, organisations or governments.

Outcomes can have short, medium and long-term achievements.

Short-term outcomes The first-order effects of the initiative, which generally include changes to participants or the community.

Medium-term outcomes The second-order effects of the initiative, which can include changes to policies, plans and projects.

Longer-term outcomes The third-order effects, or the ultimate impact that the initiative should achieve, which can include fundamental changes in the social, environmental, economic and governance priorities of the government.

The longer-term the outcome, the more likely that it will have been affected by factors external to the program that is being evaluated, and the longer you may have to wait until the outcomes are able to be assessed. Depending on the time available for your evaluation, it may only be possible to evaluate short to medium-term outcomes. However, whilst the evaluation of the longer-term outcomes is more challenging, it is also important.

Identify key stakeholders

Various stakeholders will be interested in the evaluation process and results, and should be considered in developing the evaluation model and reporting mechanisms. Exploring the various expectations of stakeholders will clarify the purpose of the evaluation.

Program partners Who has a role in developing or delivering the program? This may include central and line agencies, and non-government organisations who deliver services.

Program participants Who will participate in the evaluation? This may include community members and clients.

Evaluation audience	Who has commissioned the evaluation? What are their expectations? Who is interested in the results? This may include Ministers and Cabinet, high-level government committees, the general public or specific communities.
Research community	Is there interest in the subject of the evaluation from the research community and universities?
General interest	Is there broader interest from the general public or the media?

Form an evaluation committee

It is useful for any evaluation effort to be guided by an evaluation steering committee that is responsible for setting the overall direction and managing the evaluation, whether it is conducted internally or by external consultants. A well-selected evaluation steering committee can also guide those officers with little knowledge of evaluation through the process. To be optimally effective, membership of evaluation steering committees should be drawn from key stakeholder groups. However, depending on the nature of the evaluation and the level of confidentiality, it may not be possible to include all relevant parties.

The work of the evaluation steering committee may involve:

- determining the purpose of the evaluation;
- making decisions about the research design and elements of the evaluation;
- managing data collection, storage, and analysis;
- appointing and managing external consultants;
- interpreting results, particularly how they relate to policy development;
- monitoring progress toward program outcomes; and
- approving evaluation reports for publication.

Identify key timeframes

The timeframes for completing the evaluation will impact upon the questions that can be answered and the methods that are appropriate. It is also important to allocate sufficient time to plan the evaluation, collect the data, analyse the data and prepare the final report. Depending on the nature of the report, it is often important to circulate drafts for review, a process that often takes a significant period of time.

Determine key evaluation considerations

The assessment of program characteristics, the interests of key stakeholders, and available timeframes provides vital information to the selection of an appropriate evaluation model from the extensive range of approaches. Drawing on this knowledge to inform the specification, in turn, of the evaluation

function, type, research questions, methods and analyses will ensure a targeted and meaningful evaluation.

You may need to revisit your decisions about evaluation function and type, research questions, method and data analyses as your evaluation planning progresses. Remaining flexible between choices allows for the particular constraints of the study, such as a lack of time or resources, to be accommodated.

Evaluation function

An evaluation may be undertaken for a number of reasons. Consider what functions your evaluation is intended to perform. These functions may include:

- clarifying the goals and process of the initiative;
- monitoring outcomes;
- demonstrating accountability;
- further developing the initiative;
- demonstrating the value of the initiative; and
- promoting awareness of the initiative.

Evaluation type

Depending on the functions of your evaluation, different types of evaluation are more or less relevant in different circumstances. The program evaluation hierarchy in Figure 3 outlines the progression of evaluation types, and the issues and questions which are the focus of an evaluation.[6] It may be helpful to think of the different evaluation types as 'building blocks', where each rests on those below. However, in any particular evaluation, it may be necessary to combine more than one type to achieve the major purposes of the study.

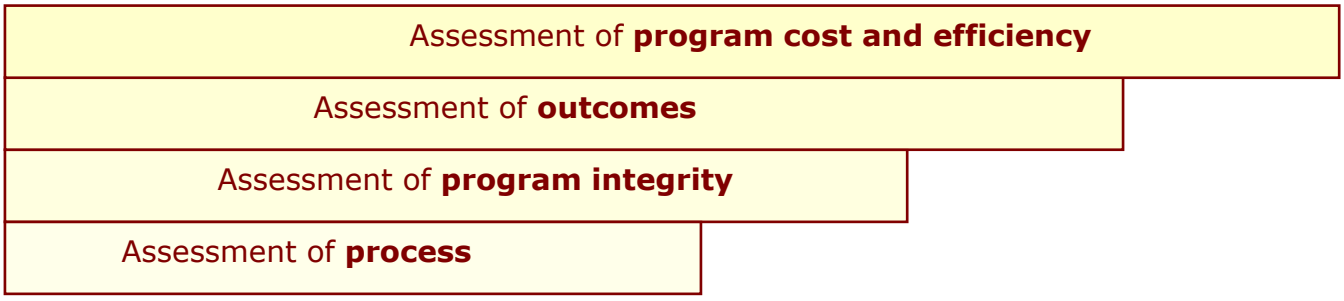

Figure 3. Hierarchy of evaluation types

The foundation block in the evaluation hierarchy is to assess the nature and quality of program implementation – the task of process or program integrity evaluation. It is only once we've effectively implemented program activities and services that it may then be meaningful to assess program outcomes. At the top of the hierarchy we have the assessment of program cost and efficiency. Asking

cost evaluation questions generally assumes that there is information available about the nature of the program, and the effectiveness of its implementation and program outcomes.

When developing the questions for an evaluation, therefore, it may be helpful for you to start at the bottom of the hierarchy and consider what is known, and what still needs to be known, before moving to the next stage in the evaluation hierarchy.

Process

Process evaluations provide information about the implementation, operation or delivery of a program with a view to improvement. They are concerned with what actually happens in practice, and typically focus on the model of service delivery. This type of evaluation may investigate how well resources are invested to achieve specified objectives.

Process evaluations aim to answer some of these questions:

- What are the key program components (i.e., activities, events, practices)? Are these effective?
- What service delivery models are used? Are these operating according to plan?
- Who participates in which activities? Are there any barriers to access?
- How well is program demand met?
- What resources and inputs are invested? Is this level of investment adequate?
- How is implementation progressing at different sites?
- What stakeholders are involved and what roles do they play? How could these roles be improved?

Program integrity

Evaluation is often conducted to monitor the fidelity of program practice with the original program design or policy. Evaluation can help to verify that the specified target populations are receiving the promised services, or the procedures are being carried out as indicated.

Importantly, successful process evaluations must not solely be defined in terms of whether or how well the desired outcomes are achieved. Numerous factors, many of which are not considered in the program design, influence whether or not an initiative achieves its objectives. Therefore, it is possible for a program to be implemented with complete integrity, but still fail to meet its aims because of the impact of unexpected factors.

Often program integrity evaluations form part of a broader process evaluation. Consequently, many of the questions asked in process evaluations are relevant in this context. Additionally, program integrity evaluations aim to answer some of the following:

- Are staff implementing the program as planned? If not, what areas are most vulnerable to variation and why?
- Are there any expected or unexpected factors negatively or positively impacting upon the program? How can these be accounted for in the future?

Outcomes

Evaluation is also conducted to determine the degree to which the desired program outcomes have been achieved, or the interim progress that has been made toward their achievement. This type of evaluation is powerful in arguing for program continuation, as it provides evidence that a particular program is related, or has contributed, to a positive change. However, it is dangerous to attribute positive outcomes to the initiative alone. Many other factors, external to the initiative, can affect outcomes, and evaluation provides merely an indication that an initiative is successful.

Program effectiveness can be measured in terms of its short, medium, or long-term impact:

- **short-term outcomes** (first-order effects) include changes to participants or the community as a result of the program;
- **medium-term outcomes** (second-order effects) include changes to policies, plans and projects as a result of changes to participants or the community; and
- **long-term outcomes** (third-order effects) include the fundamental changes in the social, environmental, economic and governance priorities of the government.

Outcome evaluations aim to answer some of these questions:

- Who benefits from the program?
- What do these people do differently as a result of the program? Are these behavioural changes sustained over time?
- Are participants satisfied with what they gain from the program?
- Which program activities or components contribute most or least to short and long-term outcomes?
- What factors mediate successful program outcomes?
- What are the social, economic, and environmental impacts (positive and negative) on people, communities, and the environment?
- What, if any, are unintended secondary or negative effects?

Cost and efficiency

Assessments of cost and efficiency aim to answer questions about program costs in comparison to either program benefits or the program's effectiveness in terms of the changes it may have addressed.

The approach that is selected to measure the relative costs, benefits and effectiveness of criminal justice initiatives can have an important effect on future spending priorities and decisions about the direction of initiatives. The terms 'cost-benefit analysis' and 'cost-effectiveness analysis', along with many others, are frequently used interchangeably. However, whilst they measure similar things, they provide different and separately useful information. Further distinction is provided later in the CJEF (refer to Financial or economic analysis).

An important component of criminal justice cost and efficiency analyses is the identification of the costs and benefits associated with the social impacts of crime. Whilst costs are measured in dollar terms, benefits are frequently expressed in physical units (e.g., number of offenders diverted from custody). Relative cost-effectiveness evaluations such as these are rare, although they seem to be increasingly recognised as needed within Government.[7]

You may find the Cost-benefit Analysis Guidelines, produced as part of Queensland Treasury's Project Assurance Framework, helpful when designing your cost and efficiency evaluations.[8]

In addition to the process and outcome questions outlined above, cost-benefit evaluations aim to answer some of the following:

- Are the program's accomplishments worth the resources invested?
- How does the cost of the program compare to alternatives?
- Which activities are the most cost effective? Least?
- How efficiently are clientele and agency resources being used?

Research questions, method and analysis

In reality, moving from a general idea about the evaluation type to specific research questions, and in turn an evaluation method and appropriate analyses, involves balancing your needs with key factors in the research and operating environments. These factors will also impact upon each other. Evaluators must make trade-offs to develop a design that best caters to factors critical to ensuring the success and relevance of the evaluation. In making these trade-offs, it is essential that key research questions are not compromised. Figure 4 outlines these factors, and the impact of these factors on the research design is discussed below.

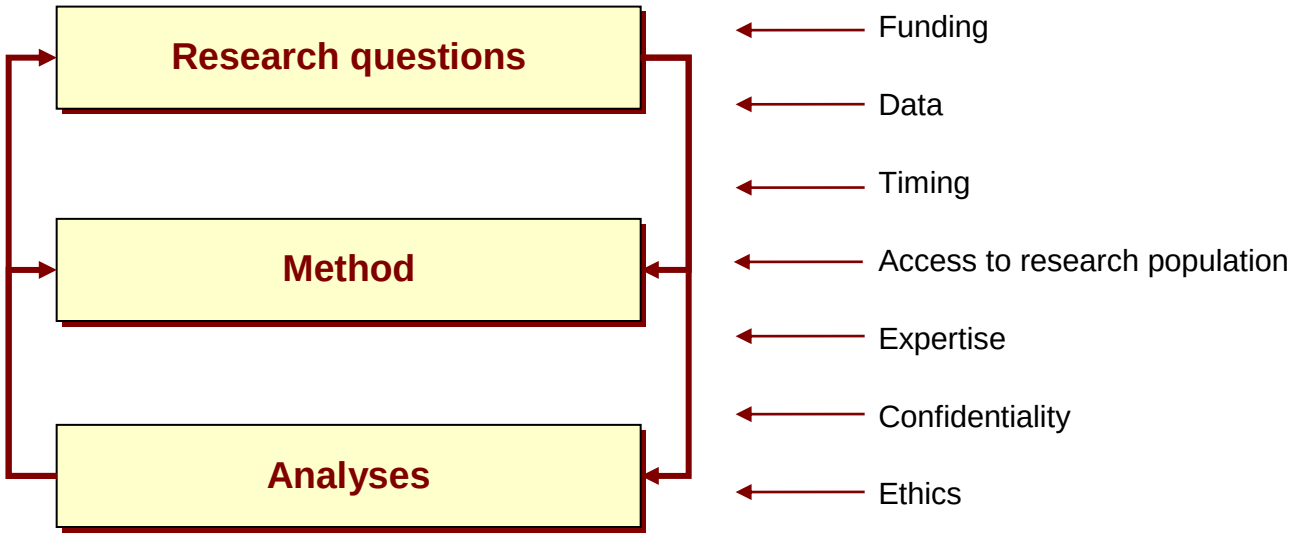

Figure 4: Factors impacting on the development of research questions and method

Funding

The funding for the project will impact upon the scale of the research, the data that can be collected, and whether an external consultant can be commissioned to undertake the research.

Data

Certain research questions can only be answered by conducting specific analyses, and the nature of the data available will heavily influence the types of analyses that can be conducted. Some analyses require data to be coded in a certain way, and when you are accessing existing datasets, it may be difficult or time-prohibitive to modify this coding. Common quantitative data analyses used in evaluations are contained in Appendix B.

Where it is important that you conduct a certain analysis to fully answer your research question, and appropriate data is not already available, collecting data may be an option. Although data collection for this purpose is often necessary, it can be costly and take significant time.

Timing

If the report is required quickly, it may not be possible to collect data. Consequently, if your timelines do not allow for data collection, evaluators may have to slightly alter their research questions and use methods that rely on existing data.

External consultants can often conduct research and produce a report more quickly than government officers that have competing interests.

Access to research population

The level and ease of access to the research population is another critical factor. For some quantitative analyses, a certain sample size is necessary in order to achieve statistical reliability. Similarly, some qualitative methods require that evaluators spend significant time with research populations. Some populations (e.g., Indigenous people, young people) are highly researched, so it is important to refine your inquiry to minimise intrusion.

Even accessing case files or official records can be difficult, so evaluators should determine whether the required data is available for analysis prior to finalising research questions and methods. It is important not to dismiss any key research questions due to availability or access problems, however.

Expertise

Some analyses are relatively simple, however, when you require more advanced analyses a consultant may be required. Consultants often have significant expertise in designing and conducting evaluations.

Confidentiality

Your selection of data collection methods will need to consider the level of anonymity and confidentiality that participants and respondents require. This is largely dependent on the nature of their participation, and the program itself.

Confidentiality and privileged communication safeguards must be included in any contractual arrangements with external evaluators.

Ethics

Regardless of whether the research is conducted internally or externally, all decisions and actions taken by evaluators whilst planning, conducting and reporting on the evaluation must be ethical.

Refer to the CJEF's later discussion, the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Research Involving Humans and Guidelines for Indigenous Research, the Australian Evaluation Society's Guidelines for the Ethical Conduct of Evaluations, and internal agency units[9] for further guidance on ethical decision-making.

Devise research questions

Whilst there are a number of factors that inform the feasibility of research questions, evaluators must ensure that it is the research questions that are driving the evaluation. Research questions must be devised prior to consideration of research methods and analyses. As planning progresses, the research questions can then be revisited. If a particular research question is essential to the success of an evaluation, however, do not allow funding or access issues, for example, to be major barriers to its inclusion. It is advisable that you explore alternative data methods, or simply alter your research questions, as opposed to disregarding them entirely.

Consider the evaluation type when developing specific research questions. Engaging key stakeholders or the evaluation steering committee in identifying research questions helps ensure the selection is meaningful. A high quality research question will:

- reference aspects of the program type and specifications;
- be unbiased;
- be ethical;
- have theoretical or practical significance;
- be achievable, given your time and resource constraints;
- not be too general, resulting in a multitude of sub-questions; and
- not be too narrow, ruling out the emergence of other possibilities.

In devising research questions, it is important to define key variables. A variable is a measurable or observable characteristic that may assume more than one of a set of values. Variables can be used to indicate whether a particular objective has been met. There are variables of process and program integrity (applied to inputs and outputs), and outcome variables.

Determine evaluation methods

A wide range of research methods, including the various alternative research designs and techniques of data collection and analysis, may be applied to criminal justice initiative evaluations. Once again, it is vital to select an evaluation method that will help you to achieve your evaluation purpose. An appropriate method should also reflect the program characteristics and the type of evaluation required.

The five broad evaluation methods discussed below reflect research designs which might be used to conduct an evaluation. Either qualitative or quantitative data, or both, might be collected within any of these designs. Appendix C provides further information on this distinction. The decision to collect qualitative or quantitative data should be made once a design choice has been made. When evaluating a policy or program, it is highly likely that you will use a combination of methods.[10]

Triangulation is the research term used to refer to the combination of different data sources (for example, data produced using different techniques, in different geographical locations, or at different points in time) and methods to provide more complete, or valid findings. Essentially, triangulation harnesses the fact that the inherent strengths and limitations of different measurement tools and methodologies complement and counteract each other, a combination of which may result in more robust and generalisable findings. By bringing together data from different sources, a comprehensive, representative and holistic assessment can be made of complex research questions. Additionally, the use of a variety of methods to evaluate an initiative allows for the accurate representation of different dimensions of initiatives, the pursuance of multiple stakeholders' interests, and the explanation of any diverse findings. Practically, however, the use of multi-methods may place high demands upon time and resources.

Particular evaluation methods are useful for achieving certain purposes. In determining the effect of an intervention on the incidence of reoffending, for example, obtaining quantitative data using an experimental or quasi-experimental design may be the most appropriate. However, ethical, access or other constraints may make it impossible or unsuitable for you to use such methods. Similarly, the use of naturalistic methods to obtain qualitative data may be useful for process evaluation, but the extensive nature of the program or resource constraints may mean that an ex post facto design using quantitative data is the most feasible method.

When deciding which type of information to use for your evaluation, consider the purpose, context and audience for the evaluation:

- Does the evaluation question lend itself to quantitative or qualitative data or both?
- What type of information is the audience most likely to expect, understand and consider credible? Are they likely to be more receptive to statistics-based evidence or illustrative case studies?
- What types of information are the participants most likely to provide willingly?

Experimental methods

A true experiment assesses an initiative by comparing initially equivalent program (individuals who participated in the initiative) and control (individuals who did not participate in the program, but share characteristics with those who do) groups. It is even possible to divide, and further compare, the program group into completers (individuals who completed the initiative) and non-completers (individuals who started the program but did not finish).

However, for a number of reasons it is rare for experimental methods to be fully and appropriately applied in public sector program evaluation. Firstly, experimental methods require a high degree of control over how an intervention is administered. This level of control is difficult or even impossible to achieve in most program evaluations. In addition, experimental designs require participants to be randomly allocated to groups, and this is frequently limited by ethical and access constraints.

Quasi-experimental methods

In many cases, the randomisation of program participants is not possible and, consequently, there can be no equivalent experimental and control groups in the true experimental sense. Quasi-experimental methods are able to adjust to the constraints of the program. In these research designs, individuals or groups are selected to serve as ‘quasi-controls’. The control group is often matched to the program group on similar characteristics. The design format is retained from the experimental model, where both the program and control groups are measured pre and post implementation.

Because of the lack of strictly equivalent groups, any observed differences may be attributed to a variety of explanations. This makes the interpretation of the findings, including conclusions about the efficacy of the program, somewhat uncertain in quasi-experimental methods. One way to improve interpretation is

to make the comparison group as similar to the program group as possible.

Survey research methods

Survey methods tend to be more descriptive, relying on the reports of participants and other stakeholders. This often makes these methods particularly appropriate for obtaining information on perceptions of a program's context, processes and outcomes.

Survey research methods are characterised by a strict adherence to formal sampling designs and a commitment to obtaining high response rates so as to ensure a representative sample of respondents. They can be particularly useful in identifying the various perspectives held on a program and its effects. Surveys may be difficult to implement well in an evaluation context, however, due to problems in developing an appropriate sampling frame and differential access to sub-groups within the sample – such as current and previous program participants and non-participants. There is also a tendency towards high refusal rates and sanitised responses among those who think that their access to services or their jobs might be adversely affected by providing negative comments on the program, although this effect may be lessened somewhat by allowing for anonymous responses.

Naturalistic research methods

Naturalistic research methods can provide useful in-depth information about a program through the use of extended interviews with open-ended questions and participant and non-participant observation, allowing the detailed exploration of significant issues. These methods may, however, have problems similar to those encountered in surveys.

The use of deliberative rather than formal sampling procedures may make sampling easier, but can raise serious questions about the representativeness of the information, particularly as resource constraints will usually severely limit the range and number of sources of information which can be used. This can lead to limited credibility of the information obtained, a shortcoming which may be overcome to the extent that the information clearly contributes to and is compatible with a well-argued case about the nature and achievements of the program.

Ex post facto research methods

Many evaluations wish to examine a program that has already been operating for some time. These studies are essentially retrospective, and ex post facto methods may be employed. Case studies, for example, examine a particular case (i.e., program, group of participants, single individual, site or location) in-depth, relying on multiple sources of information to provide a complete picture. Case control studies may be used to compare those who participated in the program to those who did not.

Clearly, in these situations it may not be possible to observe all program processes and stages, or to follow up a representative sample of program clients during and after their participation. Thus, whilst useful program information can

be gathered during ex post facto studies, the range of possible explanations and conclusions must be explored.

Financial or economic analysis

Financial and economic analysis methods aim to examine program effects, costs, and efficiency. As noted earlier, many specific techniques with this aim exist. Whilst they measure similar things, they provide different and separately useful information.

Forms of financial or economic analysis include:

- A **cost (or budget) analysis** provides detailed information on funding sources and expenses to show the estimated impact of the program on agencies' budgets. There is no measurement of benefits, efficiency or effectiveness. This analysis can be used by decision-makers when identifying factors that need to be considered for replicating a program elsewhere, or for informing budget projections.
- A **cost-effectiveness analysis** determines how much is spent on a program in order to produce a particular outcome, or how much of a particular benefit will result from a given expenditure. Benefits are identified, therefore, but are not expressed in monetary values or compared directly to costs. Additionally, each benefit is analysed individually, and no attempt is made to aggregate benefits.
- A **cost-savings analysis** is restricted to the direct costs and benefits realised by a program's funding body, which is frequently a government agency. The benefits are expressed as dollars. This kind of analysis can be used by Governments to determine whether funded programs are viable and justified in financial terms.
- A **cost-benefit analysis** involves a comprehensive economic evaluation of all the costs and benefits associated with a program, including financial, environmental and social, and in terms of productivity. This approach places benefits and costs in comparable terms, usually dollars. Benefits that cannot be expressed in dollar terms cannot be compared and are included only for discussion. The objective of this analysis is to determine a program's value, and the most economic use of resources.

Determine data analyses

The aim of data analysis is to synthesise information in order to make sense out of it. During your evaluation planning, you will need to determine which statistical techniques are most appropriate to answer your research questions.

Different techniques are appropriate depending on whether you have qualitative or quantitative data. It is also important to consider early in the evaluation process how findings from your qualitative analysis will relate to any quantitative statistics.

Quantitative analyses

The simplest way to analyse quantitative data is to use statistics that you can illustrate with tables or charts. During your evaluation planning, you will need to determine which statistical techniques are most appropriate to answer your research questions.

Quantitative research questions usually represent one of three categories: descriptive, correlational or comparative.[11] Knowledge of the type of quantitative research question can help you select an appropriate statistical analysis. Descriptive questions can be answered by using **descriptive statistics**. Answering comparative and correlational questions relies on **inferential statistics** to make inferences from the data itself to more general conditions.

Descriptive statistics

Descriptive statistics summarise the basic features of the data.

Descriptive statistics aim to answer some of these questions:

- How many offenders completed the program during its first year of operation?
- What was the rate of property offences across Queensland in 2005-06?
- How diverse are attitudes towards policing among local community members?

Inferential statistics

Inferential statistics are the outcomes of statistical tests. These statistics are useful to test hypotheses and relate findings to the sample or population.

Inferential statistics are necessary to be able to answer comparative and correlational questions. For instance, you may wish to make conclusions about the population based on data collected from a sample, or determine whether an observed difference between groups is due to an intervention or simply by chance.

Inferential statistics aim to answer some of these questions:

- Do the reconviction rates of offenders who completed the program differ from the reconviction rates of offenders who failed to complete the program after a two year follow-up period?
- After controlling for prior offence history, what is the difference in conviction and sentencing outcomes between male and female offenders?

Table B.1 in Appendix B provides further examples of research questions, and how each type of question relates to common data analysis techniques. It is important to ensure that the links between your research questions and data analyses are clear.

Statistical methods make a number of assumptions concerning the data to be analysed (e.g., that the shape of the data distribution resembles a normal or bell curve). Ideally, the analytical method that you select should be **robust**, meaning that it is not overly sensitive to these assumptions being violated. During your analysis, you should assess whether the statistical assumptions have been met. It is important to document all statistical methods, assumptions and checks for robustness in your evaluation plan and final report.

Your analyses may involve one variable (and be classed as **univariate)**, or two (**bivariate**) or many (**multivariate**) variables. When selecting an appropriate quantitative analysis, it is essential to consider the number of variables that are specified in the research questions. Variables can also be classified as continuous or categorical.

Qualitative analyses

It is different analysing qualitative data, especially when the answers are narratives, such as those you would get from interviews or open-ended survey questions. Coding, the process of using labels to classify and assign meaning to pieces of information, helps you to analyse and make sense of qualitative data.[12] Coding enables you to organise large amounts of text and to discover patterns. When you come to write your evaluation report, you can then describe the patterns and illustrate them with quotes. Refer to the footnoted reference for a more detailed description of coding procedures.[13]

Software packages are also available for computer-assisted coding, indexing and searching of qualitative data. The benefits of using software include that it eases workload, saves time and generally enhances the power of qualitative analysis.[14] However, when considering whether to use software, you will need to consider whether the software and its features facilitate the analytic procedures you wish to use. Refer to the footnoted references for further advice regarding how to select a software package to match the characteristics of the data set involved.[15]

Variables and data

After the evaluation questions and data analyses have been identified, it is necessary to decide on what information you will need for each. This section of the CJEF outlines a process for selecting appropriate and reliable data for your evaluation. Whilst detailed guidance on collecting and analysing data is not provided, a list of resources is provided at the end of the CJEF if further information is needed.

Identify data sources

The first stage in deciding on your data requirements is to make a list of all the data that you would ideally select to measure the variables – your data sources. At this stage, don't worry about whether the data is available, how it is going to be collected or its format. What you want to create is an idealised list of data. The next stages of the process will help you define what you need more precisely.

First, you must consider how much detail the data needs. For example, you may need:

- broad crime data about trends on all crimes;
- data about individual crime types;
- national or state-wide information;
- information about tightly defined geographical areas, such as individual streets, locations or properties; or
- data about specific times and dates.

The most common sources of evaluative information fall into three categories: existing information, people, and observations.[16]

Text Box 2: Guide to data and detail

The level of detail you need depends on what you want to use the data for. Collecting and analysing detailed data can be expensive and time consuming. Therefore, it is essential to plan ahead and only collect as much detail as is needed for the purposes of the evaluation. Generally speaking:

- Detailed data helps to pinpoint problems and gives an accurate picture of what has happened. However, the general picture might get lost in all the detail.
- Higher level data is useful for showing general trends, but is not as useful for detailed analysis.

Existing information

If outcomes can be measured by utilising existing administrative datasets, data sharing and linking can assist agencies to conduct better evaluations. Check to see if the information you require is already available and of adequate quality. A number of existing data sources are expanded below.

- **Program documents:** enrolment reports, case management files, statistical reports, personnel records, workplans, receipts, logs, minutes of meetings, proposals, project records, grant records, media releases, newsletters.
- **Agency databases:** Department of Justice and Attorney-General, Queensland Police Service, Queensland Corrective Services, Department of Communities.
- **External databases:** Australian Bureau of Statistics, Australian Institute of Criminology.
- **Media records:** media releases, media feature stories.
- **Other evaluations** of same or similar programs.

A list of publicly available datasets from key state and national agencies and organisations that may be useful sources for criminal justice evaluations is provided in Appendix D.

People

People are the most common source of information for an evaluation. They provide this by their actions, by volunteering comments, by having their knowledge or skills tested, or by responding to questions. Think about who is able to best answer your research questions:

- program participants;
- program managers, staff, administrators and volunteers;
- people with special expertise (e.g., judges, academics);
- community members;
- victims;
- collaborators/competitors;
- funders; or
- policy-makers and legislators.

Observations

Evaluative information is also available through the direct observation of situations, behaviours, program activities and outcomes. The advantage of observation methods is that they do not depend on people's willingness and ability to furnish information. A number of subjects exist for observation:

- program events and activities (to record the numbers, characteristics, practices, interaction patterns and skill development of program participants);

- practices;
- verbal and nonverbal behaviour;
- physical surroundings; or
- relationships.

Think about what you want to know, and determine who or what can best provide that information. Next, determine whether the data you would ideally like is available from existing sources. Sometimes, evaluating the impact of criminal justice interventions requires a greater depth of information about individuals than is currently available from existing databases. Because these systems are operational, and not research-based, they include data pertaining to the progression and processing of the alleged offender through the criminal justice system. Evaluations of specialised courts of therapeutic jurisprudence, for example, may need to gather additional data relating to the underlying causes of criminal behaviour (e.g., mental health issues, drug issues, family or individual characteristics).

If existing information is not sufficient, alternative options to obtain data include:

- collecting data for individual evaluations;
- developing specific databases to enable the collection of data for long-term evaluations;
- expanding the current criminal justice databases to enhance their utility; and
- linking data between and within government agencies (e.g., health, housing, education, child services).[17]

Finally, remember that several sources usually provide a more complete and credible evaluation than just one.

Determine sampling method and size

If you are evaluating a small program, such as a workshop, and the number of participants is small, it may be appropriate to collect data from all involved. However for larger initiatives, or when resources are limited, it is often impossible to conduct evaluation with every member of the targeted population. For these evaluations, it is necessary to draw a sample from the population. Figure 5 illustrates a three step sampling process to help with your evaluation planning.

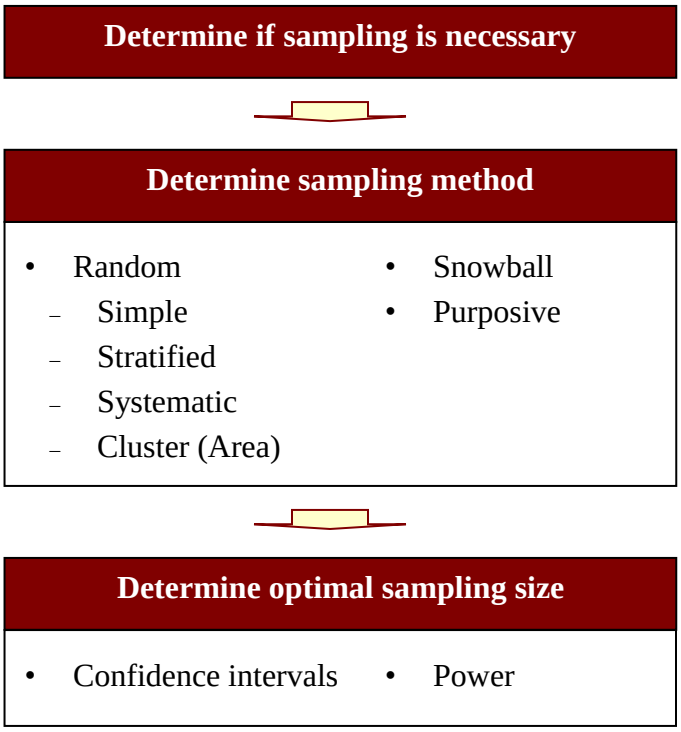

Figure 5. Three-step sampling process

Determine if sampling is necessary

First, determine if data should be collected from an entire population or from a sample. When making your decision, consider:

- the purpose of the evaluation;
- the size of the population(s);
- the method(s) of data collection; and
- the resources available to collect data.

Determine sampling method

If the evaluation sample is not proportionally representative of the population, you will not be able to generalise about the population from your evaluation results. **Sampling bias** occurs when certain population values are over or under represented in your sample. These values may 'confound' your evidence, rendering your evaluation unable to determine if a particular outcome is an effect of the initiative or an effect of a confounding variable. Therefore, your sample should reflect the demographic and other important factors which are characteristic of the population. For example, if your population is 60% Indigenous, and Indigenous status is considered an important attribute, then your sample should also be 60% Indigenous.

Certain sampling methods will help you to select representative samples and eliminate sampling bias. If everybody in the population has an equal chance of being selected, the likelihood of an unbiased sample increases. If possible, randomly select the evaluation sample from the population. Depending on the scope of your evaluation, there are several ways to do this, such as:

- using a computer program that will randomly do the selection for you;
- selecting cells on a table according to a formal, predetermined procedure; or
- pulling strips of paper out of a hat.

Random sampling strategies can be more complex and flexible than simple random selection, however.

Systematic random sampling

Quite literally, this is random sampling with a system. A starting point for sampling is chosen at random, and the population is thereafter sampled at regular intervals. For example, suppose you want to sample eight participants from a list of 120. Given that 120/8=15, every 15th participant is chosen after a random starting point between one and 15. If the random starting point is 10, then the selected participants are 10, 25, 40, 55, 70, 85, 100, and 115.

Systematic sampling is often easier to conduct than a simple random sample. It also has the benefit of spreading the sample more evenly over the population. However, be careful that your system does not interact with some hidden pattern in your population. This can be overcome by ensuring that the population is listed in a random order, at least with respect to the characteristics that you are measuring.

Cluster (area) random sampling

Sometimes is cheaper to 'cluster' the sample in some way, e.g., by selecting respondents from certain geographical areas, time-periods, or institutions only. Cluster sampling involves a three-step process:

- divide the population into clusters (e.g., along geographic boundaries, dates);
- randomly sample clusters; and
- randomly sample participants within sampled clusters, or measure all participants within those clusters.

The primary benefit of cluster sampling is that it reduces travel and other administrative costs. A geographically dispersed population can be expensive to survey, for example. Greater economy and efficiency than simple random sampling can be achieved by treating participants within a local area as a cluster.

In general, researchers prefer random sampling methods over non-random ones because they are considered representative of the population and, consequently, more accurate and rigorous. However, there may be circumstances where it is not feasible, practical or theoretically sensible to do random sampling. Some non-random sampling methods are described below.

Snowball sampling

Snowball sampling involves capitalising on informal social networks to identify respondents. For example, once having identified someone who meets the criteria for inclusion in your study, you then ask them to recommend others they may know who also meet the criteria.

Although this method would hardly lead to representative samples, there are times when it may be the best method available, for example, to reach participant populations that are inaccessible or hard to find.

Purposive sampling

Purposive sampling encompasses a range of methods that involve sampling with a purpose in mind. In all of these methods, evaluators know what they want. They might seek to sample one or more specific predefined groups or types of people, or sample for diversity.

Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern. With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.

Determine optimal sampling size

Your sample will typically not perform exactly the same on evaluation outcomes as the entire population. The larger your sample, the more reliable it is as an estimate of the population. However, the larger your sample, the more resources are required to collect data. You will need to weigh up these consequences to determine your optimal sample size.

First consider:

Variance

The more uniformity you expect in your data (or, in statistical terms, the smaller the standard deviation you expect), the less you need to be concerned that statistical error will lead you to an false conclusion that the initiative either makes a difference when in fact it doesn't (a Type I error), or doesn't make a difference when in fact it does (a Type II error).

Minimum effect size

This relates to what constitutes sufficient evidence that an initiative is making a difference in the eyes of stakeholders. An intervention is usually only deemed successful if it made a significant difference to outcomes. The effect size is how much of a difference is needed. For example, stakeholders might consider a new program successful only if it reduced substance use rates by 15% or more.

Determining variance and minimum effect size is often a matter of judgment. Use your knowledge of the field, prior research if there is any, and stakeholders' needs. The smaller the effect you expect, the more you need a large sample to conclude, with an acceptable level of confidence, that you have reliably observed a change in the population.

You can choose to use either power or confidence interval theory to determine the optimal evaluation sample size. Refer to statistics books for a more detailed description of calculating sample size using either procedure.[18]

Power Focuses on how much certainty you want to have that your statistical test will identify an effect when it exists. You need to weigh the implications of different sample sizes and different effect sizes on power.

Confidence intervals Focuses on the level of confidence you want in the sample's reliability as an estimate of the population. It can be a useful tool for weighing the trade-offs of different sample sizes in relation to the anticipated effect size and the desired level of reliability in the results.

Specify data collection

If the data that you need is not available, or is not of sufficient quality or relevance, you may need to consider organising the collection of data yourself. Decisions need to be made concerning the methods of collecting data, and how frequently this is done.

Data collection methods

Given the varied approaches to evaluation, there is no single list or categorisation of data collection methods. A list follows of some of the most common methods used: surveys or questionnaires, interviews, document reviews, observations and performance measures.

Surveys/ questionnaires

An instrument that is comprised of a series of written questions which is designed to measure participants' opinions or responses.

Surveys or questionnaires may be self or group administered and can take the following forms:

- *Written survey:* questions are presented to participants on paper.
- *Online survey:* questions are presented to participants electronically or via a website.

Interviews

A conversation involving a person or group of people where questions are asked to elicit information.

Interviews can be highly structured (as in standardised and predefined) to or free-flowing and conversational, and can take the following forms:

- *One-on-one interview:* an interviewer questions one respondent face-to-face.
- *Telephone interview:* an interviewer questions one respondent via the telephone.
- *Focus group:* a number of people are invited by a facilitator to openly

discuss opinions and responses to issues in a group setting

- *Group assessment:* information is collected through the use of group processes (e.g., the structured Delphi Technique attains opinions from a group of experts with the object of obtaining a consensus).
- *Community forums:* a public meeting to which local residents are invited to share their opinions and ideas about a particular topic or issue.
- *Submissions/testimonials:* a written or verbal statement which espouses personal views and reactions on a particular issue.

Document reviews

The analysis or summary of printed material and existing information.

Document reviews can take the following forms:

- *Expert or peer review:* the assessment of a program by a group of experts in a specific field or a review committee.
- *Literature review:* a survey or in-depth review of publications available on a topic or in a specific field.
- *Program/policy documents review:* the review of a collection of materials that encompasses the breadth and scope of the program or policy.
- *Official records review:* using existing sources of information (e.g., court, police, case management files, hospital records) to measure effectiveness.
- *Log/journal/diary review:* events are recorded over time (either factually or revealing the personal perspective of the writer) and reviewed.

Observations

Collecting information about people, events or sites by 'seeing' and 'listening' attentively, either with or without the knowledge of the people being observed. Observations may be structured or unstructured.

Performance measures

The assessment of participants' knowledge, skill or performance using established standards.

Performance measures can take the following forms:

- *Test of knowledge:* a standardised procedure for measuring participants' knowledge of a particular topic.
- *Simulated problem or situation:* the use of models or mock-ups to solicit participants' perceptions, reactions and responses.
- *Activity sampling:* keeping a record of people's work activities at set

times (e.g., at set times each day or on certain days of the week).

- *Ratings:* an appraisal of the value of something or someone (e.g., by peers, staff, supervisors, experts).

In selecting an appropriate data collection method for your evaluation, you will need to consider a number of key factors. Figure 6 outlines these factors, and the questions that these factors raise are discussed below. Consider which of these factors are most critical to ensure the success and relevance of your evaluation. It is important to ensure that you are collecting data in order to answer your research questions, not simply because it is expected. Appendix E provides a summary of the interaction between each of these factors and common data collection methods.

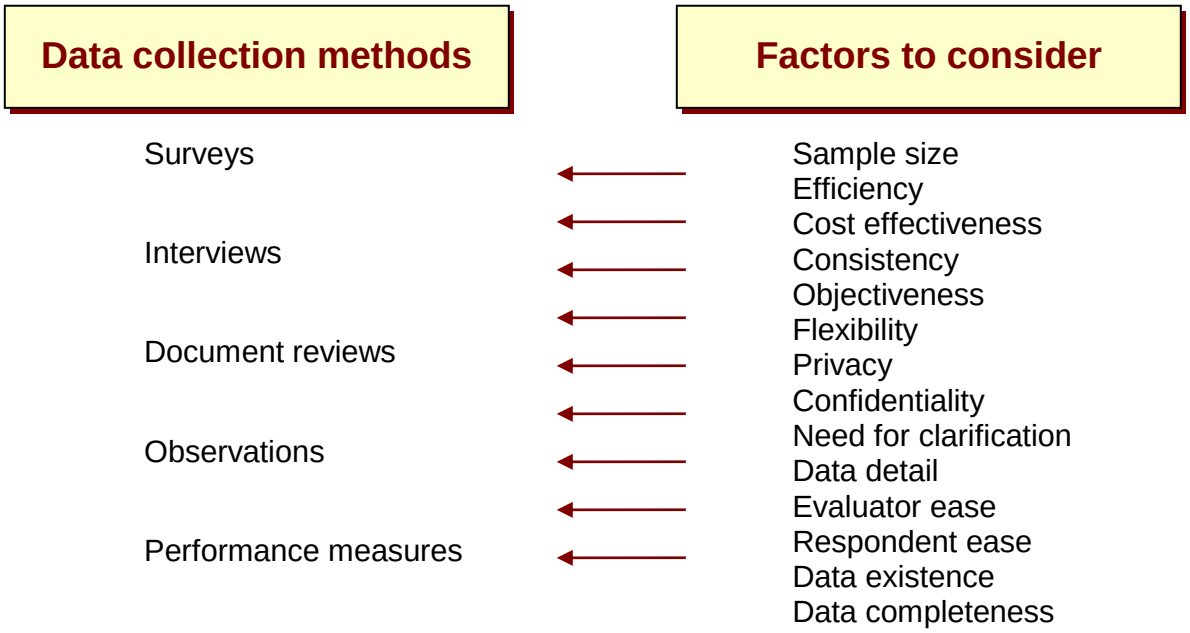

Figure 6: Factors to consider when selecting data collection methods

Sample size — What number of participants/respondents do you need to collect data from? Will you need a large sample to ensure reliability?

Efficiency — Is an appropriate amount of data produced in a minimum amount of time?

Cost — Are effective results obtained at a minimum cost?

Standardised data — Are all participants/respondents measured with the same instrument?

Objectiveness — Is it possible for the evaluator, interviewer or observer to influence responses, distort the data or interpret the data with bias?

Flexibility — Is the collection method able to be designed or altered to meet the evaluation information requirements?

Privacy — Does the collection method abide by information privacy principles (refer to Appendix F)? How will informed consent from participants be obtained?

Confidentiality — Is the anonymity and confidentiality of participants and respondents ensured?

Need for clarification — Is the participant or respondent able to seek clarification or further detail from the evaluator or interviewer when needed?

Richness of data — Does the collection method provide data that is rich in detail?

Evaluator ease — What is the burden on the evaluator (e.g., in terms of recruiting participants, analysing methods, etc.)?

Respondent ease — What is the burden on the participants or respondents (e.g., in terms of time, cost, etc.)? Are there any barriers or accessibility issues for participants (e.g., disabilities)?

How much time do potential participants have available to participate in the evaluation and is there a risk of overloading participants? Will extra support for participants be required for the more time intensive data collection activities or those that require travel, such as focus groups?

Data existence — Does the collection method rely on information that already exists?

If so, are you able to gain access to this data, and is the data in the correct format? You may need to make special arrangements to access data (refer to Develop a data sharing arrangement).

Data completeness — Has the data been recorded consistently, producing a dataset with limited gaps or missing items?

It is imperative that you consider these factors within the context of your evaluation. Data collection methods must primarily be selected based on whether they are appropriate for your evaluation purpose and questions. Ask yourself how important the collection of specific data is to the evaluation, and whether its importance justifies the method that you are using. Also consider what characteristics of your participants or respondents (e.g., age, culture, location, literacy levels, language or time available) might make different methods more or less appropriate.

Combining methods provides a valuable way to triangulate, validate your findings, and build a more thorough evaluation. Using multiple methods increases validity, but is also more costly. An inefficient or incomplete evaluation may result if you collect unnecessary data or do not match the use of methods to data needs well.

Data collection frequency

It is important that you have the data when you need it, but try and avoid being bombarded with data when it is not required. When and how often you need data depends on several factors, for example:

- Are there particular times during a program when data needs to be collected (i.e., prior to, during, following)?
- Is the data going to be used at regular intervals for monitoring the program?
- Is some data only available at certain times during a program?
- Do you need to produce reports at set intervals or when particular program milestones are reached?
- Do you need to collect data after the program has ended so that its long-term effects can be measured?

Finalise data sources

At the start of this section it was suggested that you start out with an 'ideal' list of data that you need and then refine your requirements. By this stage you will probably have found that not all data sources are available at the times or level of detail that you require, or within your budget. Your 'ideal' list has been reduced to a 'realistic' list of data.

Additionally, you should establish the accuracy and reliability of the data. Inaccurate data can distort the results of an evaluation and present a false picture of program success or failure. Whilst no data source is perfect, you should ensure that data has been accurately recorded, cleaned and analysed. Some of the questions you should ask when establishing the accuracy and reliability of any data include:

- Is the sample of population that the data was taken from representative of the target population that the program is aimed at?
- Is the data recorded correctly? (e.g., are reported offences recorded under the right crime types?)
- If any analytical packages were used, did they produce an accurate analysis of the raw data?
- Has the data been collected objectively or has the collectors' bias affected the quality?
- Is the data going to be available at the times that you require it?
- Does the data measure the same or similar things to those that you are evaluating?
- Is the data current and up to date?
- Has the data raised further questions that may need investigation?

Compare your list of data against your variables and consider whether you will have sufficient data to evaluate the program properly. If the answer is no, you will need to check whether the objectives of or terms of reference for the evaluation are still valid and the variables are feasible. Finally, it may be necessary to reconsider commissioning additional data collection exercises to collect data.

Plan for data management

Once you have finalised your data sources, you need to consider whether a data sharing arrangement needs to be made, where the evaluation data will be stored, who will be analysing the data and with what software.

Data sharing arrangement

If data sharing arrangements are relevant, then data sharing protocols need to be established to ensure that your data source remains reliable. One way to do this is to develop a Memorandum of Understanding (MOU) with each agency that you partner with. The Australian Institute of Health and Welfare has developed an example of an Australian data sharing protocol template (see Appendix F of this document) which you may find helpful.[19]

Your MOU should specify:

- the purpose of data sharing;
- what data is available;
- any services to be rendered;
- any fees to be paid;
- roles and responsibilities of partner agencies/stakeholders;
- the time frame for the partnership;
- the ownership of the data and intellectual properties;
- protection of privileged communication;
- secure storage and disposal of data;
- procedures for dealing with disputes;
- how the data will be presented in the publication and release of an evaluation report(s);
- any subsequent use of the data; and
- any relevant legislation.

A UK guide to establishing data sharing arrangements recommends that protocols be approved by senior management before any data sharing is considered.[20] In relation to client data or information that has been gathered by departmental officers, Director-General approval will need to be sought. Personal data held by agencies, or access to clients to gather such data, also needs to adhere to privacy requirements contained in the relevant legislation (refer to Privacy Issues for further information).

All partners have the right to expect that the MOU will be followed. If there is a change in circumstances, each agency has the responsibility to advise the others. Be prepared to renegotiate accordingly.

Plan for data storage

If the evaluation data are being collected, you will need to allow time for the data to be entered into a database. It is also recommended that you undertake a verification process of data entry to ensure its accuracy. Ethically, it is imperative that all information collected during the evaluation be stored, and disposed of or returned to its original source, confidentially. It is also advisable to notify participants about how the information will be recorded and stored before seeking their informed consent (for more detailed advice, refer to Consider ethical issues).

If you are using data from different stakeholders you may find that the formats are not always compatible. Some work may need to be done before data from different sources can be used together. You need to bear this in mind and allow time for any additional work to be completed. To ensure that the data is in the correct format for analysis, consider all of your data sources individually and decide:

- Can the data be used in its current format?
- Does the format need to be changed?
- Is the format compatible with the other sources of data you have?
- How much work is involved in changing the format and is it cost effective to carry out the work?

Another important step in any data processing task is to verify that your data values are correct or, at the very least, conform to some set of rules. For example, most data entered under the variable 'gender' would be expected to have only two values, 'male' or 'female'. In certain circumstances however, such as in court databases, gender may take on a third value, 'company', if the subject is an organisation and not an individual. Data from multiple sources usually needs to be transformed and cleaned before it is transported into your data warehouse. Data cleaning is the process of detecting and removing errors and inconsistencies from data in order to improve its quality. Data cleaning may involve:

- ensuring fields are labelled correctly;
- ensuring all data is formatted consistently;
- identifying missing data;
- eliminating duplicate data;
- identifying invalid or out-of-range data values; and
- identifying outliers.

Plan for data analysis

You should have already determined which statistical techniques are most appropriate to answer your research questions (see Determine data analyses above). The next step is to decide whether an external or internal evaluator will conduct the analyses.

If the analyses are to be done internally, it is necessary to ensure that adequate software is available for your chosen statistical techniques. Sometimes agencies do not have access to the appropriate software, and budgets do not allow for new purchases. If this is the case, refer to Appendix G for details on a range of free and open-source software that is available for evaluation data analysis.

Reporting strategy

Once evaluation data have been collected and analysed, and decisions about program integrity, impact, and cost and efficiency have been made, the evaluation results (or findings) should be effectively communicated to people directly involved in program implementation and to other program stakeholders. Include report writing in the evaluation plan, allowing sufficient time for necessary editing, revision and review. The evaluation will be of little benefit if the final results cannot be understood.

Identify report audience

Evaluators should revisit the intended purpose of the evaluation before drafting the evaluation report. The purpose of the evaluation dictates the audience for the report, which in turn guides its language, form, and substance. The report audience is normally the people who requested the evaluation, but you need to think about whom else should receive or be able to access the report. For example, a report for Cabinet will differ from a report for members of the public.

It is important to make evaluation findings as useful as possible to intended audiences. Consider whether those receiving the report have the skills and time to interpret findings. It is important to ensure the level of complexity, and the length, of the report is correct. For example, if you are producing a written report, it is helpful to include a summary of your main findings at the beginning so that people can decide which sections of the report they need to read. Another suggestion is to prepare reports of evaluation findings that are tailored to the information needs of various stakeholder groups.

Determine report format and structure

A detailed written report is the most common way of getting your information across, but there are different techniques you could use to report the findings and recommendations. These include: a summary report with key findings and recommendations; articles for stakeholders' newsletters; articles for newspapers; inserts in local and community newspapers; media releases; brochures, pamphlets, or other small publications; public community meetings; briefings or presentations for specific stakeholders; television or radio interviews; and website publishing. Look for common ways to inform several stakeholders and interested parties at the same time. You will also need to identify how frequently you need to report and any deadlines you will need to meet.

The report should include a description of the program being evaluated, clearly specified research questions, and an explanation of the methods and measures used to collect data to address those questions – as well as the findings arising from the effort. Appendix H provides a suggested evaluation report structure. Consider specific reporting requirements when developing the document outline and its content.

Plan for displaying data information results

It is helpful to consider the presentation of results whilst still in the evaluation planning phase. Data can be presented in the text, in a table, or pictorially as a chart, diagram or graph. Any of these options may be appropriate depending on the information that you wish to convey to the reader. Essentially, the mode of presentation that you select must match the data collected.

Well presented tables and graphs can concisely summarise information which would be difficult to describe in words alone. In general, text alone should not be used to convey more than three or four numbers. Tables are useful for providing structured numeric information, whereas graphs help to indicate trends, make broad comparisons or demonstrate relationships.

The aim of a table is to present data clearly and concisely to the reader. If you are planning to use a table to display your data, keep the following in mind:

- consider the purpose of including a table;
- ensure any descriptive information that is needed to correctly interpret the data is included;
- reference tables are usually included in an appendix;
- if the purpose of the table is to demonstrate differences between groups, the table rows should specify the different groups (because it is much easier for a reader to make comparisons within columns than rows); and
- the number of digits and decimal places used should be the minimum number needed to convey the purpose behind the table, and consistent for each variable presented.

The three main types of graphical representation of data are line, bar and pie graphs. Appendix I provides advice on when it is appropriate to use the various graphs, and some considerations to keep in mind during their development.

Ideally, tables and graphs should be self-explanatory, and able to be understood without detailed reference to the text. Despite this, the text should always include a summary of the table or figure for explanatory purposes. Graph axes, table rows and columns, and figures should be labelled clearly and informatively, with measurement units specified.

To ensure consistency, and for further advice on displaying results, it is advisable that you consult your departmental style guide, the Commonwealth Government Style manual for authors, editors and printers,[21] or another style guide.[22]

Manage review process

To prevent the potential misinterpretation of findings, allow some stakeholders to review the evaluation reports before wider dissemination occurs. Refer to your list of stakeholders and interested parties, and identify who should have the opportunity to review the information prior to release. Additionally, management often needs to peruse the document, and this can take as long as four weeks.

Evaluation management

Finalise evaluators

Another decision when managing an evaluation is whether it should be conducted by external evaluators, internal staff, or a combination of both.

- An **external evaluator** is typically an unbiased third party hired to estimate the value of a program, from a more distant perspective.
- An **internal evaluator** refers to someone conducting an evaluation from within the organisation that is implementing the program under examination.

Table 1 lists some of the benefits and limitations associated with external consultants and internal staff.

Table 1. Benefits and limitations of external and internal evaluators

Source of help	Benefits	Limitations
External consultants	• May have more expertise in evaluating programs • May evaluate more objectively • May improve the credibility of the results	• Can be quite expensive • May need extensive background briefing on the area • Less control over process (higher risk) • May not understand the context or scope of the program or the needs of the audience • Conflicts of interest may occur between meeting the needs of clients, and potentially gaining future work, and remaining unbiased • May have a desire to publish results
Internal staff	• May be more cost effective • May have ready access to existing data • Advantageous to develop the necessary evaluation skills in-house • May have experience evaluating similar programs • May increase willingness to participate	• May be biased by the evaluator's experience with the activity or desire to demonstrate certain results • Staff may lack the relevant research and evaluation skills • Management may have an influence in enhancing or politicising findings

	• May increase usefulness and uptake of evaluation results	

The flowchart in Appendix J may help you to decide whether an internal or external consultant should conduct the evaluation. When making the decision, consider:

- the available financial resources, human resources, skills and expertise;
- the time available for data collection and analysis;
- the size of the program;
- the purpose of and audience for the evaluation;
- the likely impact of the results; and
- the need for impartiality.

For some evaluations, it is vital that evaluation is a neutral process and is seen as such. There is a danger in using the same people who work on the program to evaluate it. It is very difficult for someone who has worked on a program to detach themselves from it and take an objective and impartial look at the results. Although this is unlikely to be a deliberate process, the result may be that the findings are skewed or the data interpreted in a particular way. An independent evaluator will look to see where a program has worked and where it hasn't.

Where impartiality is important, but for some reason an independent evaluator cannot undertake the evaluation, is may be useful for an external person or body to review the evaluation to act as a quality control check. Another alternative is to use a combined internal/external approach. This way, an external evaluation expert is brought in to work with agency staff to develop an evaluation design and evaluation materials, or to collect data.

You will need to assign roles for every task in the evaluation that is to be completed. For example, you need to think about:

- who is going to gather the data?
- who will analyse it?
- how will liaison between parties occur?
- who is going to write the report?

Manage tender process

When commissioning an evaluation, you will need to prepare a discussion paper or terms of reference that states the rationale, purpose and scope of the evaluation, the key questions to be addressed, any preferred approaches, issues to be taken into account, and the intended audiences for reports of the evaluation. Potential evaluators will then have the opportunity to submit a tender in response to your

discussion paper. It is important that your discussion paper strikes a balance between providing sufficient information to inform the evaluation and not being overly proscriptive, thus allowing tenderers to explore evaluation elements and be innovative in their research design.

In responding to an evaluation discussion paper, evaluators should explore the shortcomings and strengths of the brief. They should identify any likely methodological or ethical limitations of the proposed evaluation, and their possible effect upon the conduct and results of the evaluation.

You will need an agreed contractual arrangement in place between those commissioning the evaluation and the external consultants. Appendix K provides details of a comprehensive agreement. Both parties have the right to expect that the contract will be followed. If there is a change in circumstances or unforeseen conditions, each party has the responsibility to advise the other. Be prepared to renegotiate accordingly.

Consider ethical issues

Ethics refers to right and wrong in conduct.[23] This section of the framework provides some guidelines for ethical behaviour and decision making in initiative evaluation. The advice in Appendix L is adapted from the Guidelines for the Ethical Conduct of Evaluations, developed by the Australian Evaluation Society. It outlines certain procedures that you can adopt whilst planning, conducting and reporting on your evaluation which help you adhere to ethical principles.

In accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Research Involving Humans, any research or evaluation that *involves human participants* must be reviewed and approved by a Human Research Ethics Committee (HREC). These groups are established by institutions such as government departments, universities, hospitals, non-government organisations and medical facilities to provide advice on ethical issues and to approve research projects. Please note that not all forms of evaluation, desktop evaluations for example, involve human participants and hence will not be covered by the guidelines or require approval by a HREC.

Queensland Health is the only Queensland Government agency with its own HRECs. Aside from the central Queensland Health Research Ethics Committee,[24] there are a number of local committees across health service districts and hospitals. Within other Government agencies, research committees (as opposed to ethics committees) may be involved in overseeing or approving evaluation activities.

Your role in gaining ethics approval depends on whether the evaluation is being conducted internally or externally. External evaluators will often gain approval themselves, frequently from university HRECs, although this will need to be stipulated in the evaluation tender documents (refer to Manage tender process) and eventual contracts. If the evaluation is internal to government, you will need to seek advice during the planning process of your evaluation about the specific procedures for gaining ethics approval within your agency.[25]

If an ethical issue comes up during the course of your evaluation, you are encouraged to refer to the relevant ethical standards within your organisation, and discuss the issue with people experienced in

ethics or evaluation. Appendix F discusses some common ethical issues that are raised in criminal justice evaluations.

Finalise evaluation project plan

An evaluation project plan must be in place to guide the evaluation of the program. Ensure that the evaluation process flexible to allow it to adapt to changing requirements and circumstances. Depending on the scope of your evaluation, develop a project plan to document all or some of the following:

- the objectives or aims of the evaluation;
- the rationale behind the evaluation;
- the scope of the evaluation;
- the strategies the evaluation will employ;
- the benefits of conducting the evaluation;
- evaluation outcomes;
- any related evaluations, activities or projects;
- evaluation partners/stakeholders;
- details of costs and funding;
- the roles and responsibilities of those involved;
- the evaluation schedule;
- planned risk management strategies; and
- planned quality management strategies.

See below for advice on developing an evaluation budget, schedule and risk management strategies.

Before finalising an evaluation project plan, determine whether the plan can be implemented. This requires you to consider the time and resources that are available, including:

- staff numbers, available time and skills sets;
- if the data are not already available, financial resources to collect new data; and
- the time available for data collection and analysis.

While these issues need to be considered at each step in the development of an evaluation design, the potential time and financial costs of an evaluation can often not be accurately estimated until the data collection and analysis methods have been identified. If the resources are not available to collect and analyse the data in the time available, the evaluation questions may need to be revisited. In some cases the purpose or the approach to evaluation may also need to be revisited if the research design needs to be fundamentally scaled back.

Finalise evaluation budget

Managers should expect to budget approximately 10% of the overall criminal justice program budget for an effective evaluation.[26]

Although you have already listed the inputs for the evaluation, it is important to ensure that the cost of these resources does not exceed the evaluation's budget. Make a detailed list of all items that will need to be paid for from the evaluation budget. These may include:

- the cost of tendering an external evaluator;
- the cost of the evaluator;
- salary and other costs of support staff;
- travel and accommodation, meal and living expenses;
- equipment, office and support costs;
- production and distribution costs for survey instruments;
- recruitment costs for interviews and surveys;
- database development and data analysis costs;
- venue hire; and
- the production of reports and presentation materials.

Estimate the cost of each item and the total expenses. Compare the amount of money you require with the amount in the evaluation budget. Make any changes you need to keep the evaluation within budget.

Finalise evaluation schedule

Although you will have already specified a general timeframe for the evaluation, you will need to develop a detailed and realistic evaluation and reporting schedule. Appropriate times for achieving program activity outputs and outcomes will again be guided by their scope and objectives. Remember that conducting a well-planned evaluation takes time. It is important to ensure that sufficient time is scheduled to complete each evaluation task, and also allow time for review.

For example, frequently criminal justice evaluations are required to measure reoffending rates among program participants. To allow for the assessment of long-term recidivism outcomes, therefore, a viable evaluation timeframe demands a follow-up period of two years post program completion. If your evaluation does not allow for such an extensive period, you may build into the methodology the possibility for further follow-up through direct interview or via secondary data analysis.

The stages in the evaluation will take approximately the following length of times, depending on the size of the evaluation and its resources:

- Managing stakeholders: ongoing throughout the evaluation.
- Scoping the program: two to three weeks.

- Developing a draft budget and timetable: three to five days.
- Seeking approval for the budget: one to four weeks.
- Deciding upon an internal or external evaluator: four weeks.
- Appointing an evaluator: two to eight weeks.
- Evaluating the program: this will depend on the evaluation that is needed.
- Announcing findings and recommendations: one to two weeks.
- Acting on the recommendations: this will depend on the program.

Draft a timetable for the evaluation, noting:

- each activity in the evaluation process;
- how long each activity will most likely take to complete;
- how frequently you need to report and any deadlines you will need to meet; and
- the dates when events will occur and milestones will be reached.

Manage risks

Risks arise from the uncertainties which continually surround an evaluation, its operational environment and its findings. The CJEF advocates a proactive approach for dealing with evaluation uncertainty, by identifying any inherent risks in advance, and developing strategies and plans to manage them.

To determine evaluation risks, you will need to carefully assess the criminal justice program context, and the chosen evaluation form and method. Give particular consideration if the program is highly controversial, problems with the program have already been identified, or tensions exist between stakeholders.

Appendix M provides examples of some common potential risks for government evaluations. In particular, it is important to consider:

- expectations that the findings should or should not be publicly available;
- how the evaluation might raise community expectations;
- what variables are politically, culturally and socially appropriate;
- the source of funding for an evaluation;
- the timing and scheduling of evaluation outputs;
- the sources of reliable data;
- the time people or agencies may have available to participate in an evaluation;
- the willingness of people or agencies to participate in an evaluation; and

- management or political interference with findings.

However, the identification of risks should not be an excuse not to evaluate. Evaluations can almost always be modified to avoid potential risks, and in response to their occurrence. In addition to preventive management, contingency strategies are necessary for all risks.

Preventive risk management strategies aim to anticipate and avert problems before they occur. They involve performing certain actions ahead of time to either prevent a risk from occurring altogether, or to reduce the impact or consequences to an acceptable level. **Contingent** risk management strategies involve creating fallback plans that can be activated when efforts to prevent a risk fail. They address what to do if the risk occurs, and focus on how to minimise its impact. Some examples of appropriate preventive and contingent strategies are provided in Appendix M.

The continual changes that occur over the course of an evaluation means that you are required to regularly re-assess the status of known risks, and to update management plans accordingly. Additionally, you should constantly be looking for the emergence of new evaluation risks.

Review evaluation planning

Regardless of the scope of your evaluation, it is essential that you ensure it is conducted ethically and cost effectively, and that the results obtained are accurate and useful to stakeholders. The Program Evaluation Standards are principles which aim to guide the design, conduct and assessment of program evaluations. The standards address the utility, feasibility, propriety, and accuracy of evaluations. You may wish to refer to the Standards, presented in Appendix N, when planning an evaluation.

Once you have planned your evaluation and its management, the evaluation plan should finally be checked against the evaluation plan checklist in Appendix O.

Finally, for significant evaluation proposals, an Evaluation Plan should be prepared prior to Cabinet or CBRC consideration and attached to the submission for endorsement. Agency CLLOs and Portfolio Contact Officers in Policy Division of the Department of the Premier and Cabinet will be able to provide guidance as to whether particular submissions should be accompanied by an Evaluation Plan. A recommended Evaluation Plan template for Cabinet and CBRC submissions is provided in Appendix P.

Glossary

This part of the CJEF should help clarify the often bewildering range of terms and phrases used in the process of evaluation.

Categorical variable — A variable measured on a nominal scale, whose categories identify class or group membership (such as gender with classes male, female, or company).

Coding — The process of translating raw data into meaningful categories for the purpose of data analysis, and to identify recurring themes and ideas.

Confounding variable — A variable that may affect the outcome you want to examine but is not of interest for the present evaluation.

Continuous variable — A variable measured on a continuous scale (such as time in years).

Control group — A group that is not subjected to an initiative or program so that it may be compared with the experimental group who receive the intervention.

Data — Information including facts, concepts or instructions, represented in a formalised manner, that is suitable for communication, interpretation or processing.

Data analysis — Systematically identifying patterns in information and deciding how to organise, classify, interrelate, compare and display it.

Data sources — Documents, people and observations that provide information for the evaluation.

Effectiveness — The degree to which an initiative or program yields the desired outcome.

Evaluation — The systematic collection and analysis of information to make judgments, usually about the effectiveness, efficiency and appropriateness of a program or initiative.

Goal — A goal is a simple statement which sets out the purpose of a program or evaluation. It's important not to confuse goals with objectives. An objective is a specific statement that can be measured.

Informed consent — When the evaluator provides information to participants as to the general purpose of the study, how their responses will be used, and any possible consequences of participating prior to their involvement in the study. Participants typically sign a form stating that they have been provided with this information and agree to participate in the study.

Initiative — Any set of programs, procedures, activities, resources, policies or strategies that aim to achieve common goals or objectives.

Input	The inputs to a program are the resources used to carry out the work. Resources can be financial, material or human. From a program management point of view it is important to be aware of exactly what resources are available to carry out the work. When resources are limited, it can affect the objectives of a program and the scope of the work carried out.
Instrument	A tool or device (e.g., survey, interview protocol), used for the purpose of evaluation.
Objectives	An objective is a statement that describes something you want to achieve – the desired outcome of a program or an evaluation study. It is important that objectives are written so that they are **SMART** (**s**pecific, **m**easurable, **a**chievable, **r**ealistic, **t**imebound).
Outcome	The overall result of applying the inputs and achieving the outputs. In other words, the effect or change resulting from an initiative or program.
Outlier	An unusual value that is correctly reported but is not typical of the rest of the population.
Output	An output is a piece of work produced for a program. It is important to realise that an output is not necessarily the final purpose of a program. Outputs are usually things that need to be done in order to produce the desired result.
Population	The complete set of individuals or groups about which information is required. A population may share a common set of characteristics.
Program assumptions	Program assumptions are the beliefs we have about the program, the participants, and the way we expect the program to operate.
Program group	A group that receives a treatment or an intervention, or participates in an initiative or program, in an evaluation.
Qualitative variables	Qualitative variables are those that measure information in non-numeric form or qualities, which are usually quite intangible things, such as the opinions, perceptions, feelings and beliefs of individuals and groups. For example, changes in the level of fear of crime in the elderly would be a qualitative variable.
Quantitative variables	Quantitative variables measure information in numeric form, or tangible things, such as the number of burglaries, or percentage of homes burgled, in an area.
Reliability	The extent to which a measure, instrument or observer would produce the same results, if repeated using another statistically equivalent sample and methodology.

Sample	A defined subset of the population that is chosen to participate in the evaluation based on its ability to provide information, its representativeness of the population under study, and/or factors related to the feasibility of data gathering such as cost, time, participant accessibility, or other logistical concerns.
Sampling bias	Sampling bias occurs when the evaluation design fails to capture the true population and implementation characteristics, thus rendering the results un-generalisable.
Type I error	When a statistical test falsely detects an effect that does not really exist.
Type II error	When a statistical test fails to detect an effect that really exists.
Validity	The degree to which the theory and information gathered support the proposed uses and interpretations of a measure or an instrument.
Variable	A variable is a measurable or observable characteristic that may assume more than one of a set of values.

Appendix A: Conceptualising an evaluation

The logic model is a diagram which captures information about the main elements of a program and describes in concise terms how the program operates, as well as the outcomes or results that the program is intended to produce. It is a means to an end, rather than an end unto itself. It is a powerful conceptual tool that serves as the foundation for the subsequent steps of the evaluation planning. Evaluation questions also correspond with the components of the logic model.

Figure A.1 provides an example of a logic model. Developing a logic model requires the identification of six key elements: program goals and objectives, environment factors, and assumptions; and evaluation inputs, outputs, and outcomes (refer to Identify program characteristics).

As the evaluation process unfolds, the logic model is a living document that will most likely change as it is used. Program managers should regularly hold their implementation up against the logic model, to assess the degree to which it alights with program implementation. It may be necessary to adjust the logic model to reflect changes in consensus understanding about the underlying theories driving program design and implementation. It is also useful to overlay information about data collection and analysis onto the logic model, to ascertain the degree to which a complete evaluation picture is developing.

Program goals and objectives

- What is the problem that the program aims to address?
- Who are the stakeholders? Do they support the project? Will they support the evaluation?
- What are the intended outcomes of the program?
- How will the evaluation be used?

Program environment factors

- Conditions existing prior to the program (e.g., the characteristics of the community involved)
- Non-governmental factors (e.g., level of media and community interest, non-governmental concurrent activities)
- Internal agency factors (e.g., level of commitment to program within agency, capacity for program, decision making processes)
- Whole-of-government factors (e.g., whether the program is part of a broader policy or strategy, other significant programs occurring at the same time).

Program assumptions

Principles, beliefs and ideas about:

- The problem or situation
- The resources or staff (e.g., secure funding, necessary skills and abilities of staff)
- The way the program will operate (e.g., level of intervention received)
- What the program expects to achieve (e.g., prior evidence of effective strategies)
- The knowledge base (e.g., theories, best practices)
- The external environment (e.g., level of exposure to external initiatives)
- The internal environment
- The participants (e.g., how do they learn, their behaviour, motivations)

Inputs

Resources: what you invest

- Staff
- Intellectual resources
- Physical resources
- Budget
- Local service providers
- Reference group
- Policy makers
- Service delivery partners

Outputs

Activities: what you do
Participation: who does what

- Staff
- Information provision
- Consultation opportunities
- Active participation
- Audience

Outcomes (Impact)

Short-term: what happens as a result

Changes to participants':

- Perceptions and attitudes
- Awareness, knowledge and competencies
- Skills, abilities and capacities
- Behaviours and actions

Medium-term: what this leads to

Changes to:

- Policies
- Plans
- Projects
- Services
- Partnerships, alliances, coalitions or networks

Long-term: what this contributes to

Fundamental changes to:

- Social priorities
- Environmental priorities
- Economic priorities
- Governance priorities
- Participants' quality of life

Figure A.1. Logic model: The conceptualisation of an evaluation

Appendix B: Quantitative data analyses

Table B.1 displays the relationship between the distinct types of research questions and the various analyses that are available. Descriptive statistics are used to answer descriptive questions, whilst inferential statistics answer comparative or correlational questions. Whilst the table suggests a number of data analysis methods, this list is not intended to be exhaustive.

Table B.1. Examples of common quantitative analyses

Type of research question	Research question examples	Variable type	No. of variables	Suggested analyses
Descriptive questions – Seek to quantify responses on 1 variable at a time – Often begin with the words *what is…* or *what are…*	*What is the average/middle of/most common (variable)?* – What is the crime rate in a particular community? – On average, how many offenders that completed the substance abuse treatment recidivated?	Continuous or categorical	1	Measures of central tendency (e.g., mean, median, mode)
	What is the range/spread of (variable)? – What is the range of sentencing lengths for sexual assault convictions? – How diverse are satisfaction levels among community members?	Continuous	1	Measures of variability/dispersion (e.g., range, standard deviation)
	What is the relative position of a particular case within (variable)? – How does the Queensland initiative compare to others Australia wide?	Continuous or categorical	1	Measures of position/location (e.g., percentile rank, *z*-score)
Comparative questions – Seek to compare 2 or more groups on some outcome variable – Often use words such as *differ* or *compare* – Can be causal in nature by implicitly comparing 2 groups (e.g., *what is the effect…*) – Can assess changes in some outcome variable over time – Can assess differences in some outcome variable between geographical areas	*What is the difference in (outcome) between (group 1) and (group 2)?* – What is the effect of crime prevention techniques on the crime rate? – Do the offenders selected for an anger management intervention differ significantly on key characteristics from the general criminal population?	Categorical groups, continuous outcome	2	t-test
	What is the difference in (outcome) among (groups)? – What is the difference in substance use rates among offenders receiving the three different interventions that are offered? – What is the difference in crime rates among those communities that implemented the crime prevention techniques and those that did not?	Categorical groups, continuous outcome	2 or more	Analysis of Variance (ANOVA)
	After controlling for (mediator), what is the difference in (outcome) among (groups)? – After controlling for poverty rates, how do crime rates differ among communities?	Categorical groups, continuous mediator and outcome	3 or more	Analysis of Covariance (ANCOVA)

	What is the difference in (outcome) and (outcome) among (groups)? – How do sexual reoffending rates and general reoffending rates differ among different classes of sex offenders?	Categorical groups, continuous outcomes	3 or more	Multiple Analysis of Variance (MANOVA)
	After controlling for (mediator), what is the difference in (outcome) and (outcome) among (groups)? – After controlling for prior offending history, what is the difference in conviction and sentencing outcomes between males and females?	Categorical groups, continuous mediator and outcomes	4 or more	Multiple Analysis of Covariance (MANCOVA)
	What is the difference in (outcome) over (time)? – How does the crime rate change over time? – Are the substance use rates of offenders altered over the course of treatment?	Continuous	2 or more	Time series analysis
	What is the difference in (outcome) among (contiguous geographical areas)? – Does the crime rate differ among rural communities?	Categorical areas, continuous outcome	2 or more	Spatial data analysis
Correlational questions – Concerned with trends between or among 2 or more variables – Often use words such as relate, relationship, association or trend	*What is the relationship between (variable) and (variable)?* – Does male sex offending decrease as offenders age?	Continuous	2	Correlation
	What is the relationship between (variables) and (outcome)? – Does parental criminality increase a child's involvement in crime?	Categorical or continuous (depending on type of regression analysis)	3 or more	Regression
	Which (variables) discriminate among (groups)? – What distinguishes those offenders who successfully complete the intervention and those that do not? – Which personal characteristics are the best predictors of reoffending?	Continuous variables, categorical groups	3 or more	Discriminant analysis

Appendix C: Qualitative and quantitative data

Variables can be quantitative or qualitative. Some of the differences between these types of variables, in terms of uses, benefits and limitations, are described in Table C.1.

Table C.1. Differences between quantitative and qualitative data

	Quantitative	Qualitative
Description	– Numerical data	– Non-numerical data (i.e., words, text, photographs, movies, sound recordings, etc)
Uses	– To provide snapshots of numerical data that can be compared against milestones or targets – To provide evidence of trends and patterns – To examine two or more variables through statistical analysis – To identify cause and effect relationships	– To gather perceptions of events and other experiences – To explore and compare the range of perspectives – To identify important factors or variables when these are poorly understood – To provide greater detail and meaning for quantitative data
Benefits	– Increased rigour (reliability) and standardisation of results for accurate reporting	– More detailed, 'rich' information – Much broader scope with a focus on both intended and unintended outcomes – People may relate better to the results because they are often in the form of 'stories' rather than numbers
Limitations	– Generally does not provide information on unintended outcomes	– Data collection and analysis is often more time and thus resource intensive – May not be seen as equally credible, reliable and robust compared to quantitative-based methods – Often difficult to generalise findings to a large population

Appendix D: Existing datasets for criminal justice evaluations

Table D.1 provides a list of existing administrative and survey datasets from key state and national agencies and research centres. All of those listed here are publicly available, and may be useful sources for criminal justice evaluations.

Table D.1. Key national and state criminal justice datasets

Name	Description	Associated organisations	Available data	
National				
Australian Bureau of Statistics (ABS)	Collects and disseminates state justice system administrative data, population data and victims of crime surveys.	– National Centre for Crime and Justice Statistics (NCCJS)	Administrative data: – Recorded Crime – Criminal Courts – Corrective Services – Prisoners in Australia	Survey data: – Crime and Safety Survey – General Social Survey – National Aboriginal and Torres Strait Islander Social Survey – Population by Age and Sex
Australian Institute of Criminology (AIC)	National research centre for the analysis and dissemination of criminological data and information.	– Campbell Collaboration's Crime & Justice Coordinating Group	National monitoring projects: – National Homicide Monitoring Program (NHMP) – National Firearms Monitoring Program (NFMP) – National Armed Robbery Monitoring Program (NARMP) – Drug Use Monitoring in Australia (DUMA) – National Deaths in Custody Program (NDICP) – National Police Custody Survey – Juveniles Justice in Australia	Justice research projects: – Violent Crime – Environmental Crime – Organised and Transnational crime – Economic Crime – Drug and Alcohol – Property Crime – Cybercrime
Australian Institute of Health and Welfare (AIHW)	Collects and disseminates national health and welfare statistics.	– National Drug & Alcohol Research Centre (NDARC)	Administrative data: – Child health, development and wellbeing – Child Protection Data Set – NMDS for Alcohol and Other Drug Treatment Services (AODTS)	Survey data: – National Drug Strategy Household Survey – Illicit Drug Reporting System (IDRS)

Australian Crime Commission (ACC)	Reports on national and state/territory law enforcement illicit drug data.		– Australian Illicit Drug Report (AIDR) – Organised crime in Australia
Australian Institute of Family Studies (AIFS)	Central agency for policy relevant family research and data in Australia.	– National Child Protection Clearinghouse – Sexual violence research	– Australian Temperament Project (longitudinal study) – Beyond 18: The longitudinal study of leaving care – Knowledge Circle – Keeping Indigenous children safe and happy
Productivity Commission	Reports on state and national monitoring data across government including the justice sector.	– Steering Committee for the Review of Government Service Provision (SCRGSP)	– Overcoming Indigenous Disadvantage – Indigenous Expenditure Report Within ROGS – Police – Courts – Corrective services – Juvenile justice
National Legal Aid	Compiles administrative data from state and territory legal aid services.		– Legal aid services
SA Office of Crime Statistics and Research (OCSAR)	Monitors SA crime data, maintains the national database on vehicle thefts, and conducts research and evaluation.	– National Motor Vehicle Theft Reduction Council	– Vehicle theft – Justice Data Portal
State			
Queensland Police Service (QPS)	Produces the latest publicly available crime statistics		– QPS Crime Map – Reported crime trend data
Department of Justice and the Attorney-General (JAG)	Produces an annual report which provides Queensland court data.	– Queensland Courts	– DJAG annual report – Queensland Government Open Data
Queensland Corrective Services (QCS)	Produces an annual report which provides Queensland prisoner data.		– DJAG annual report – Annual reports – pre 2013 – Queensland Parole Boards Annual Report
Queensland Government Statistician's Office (QGSO)	Provides statistics relating to criminal activity, criminal justice and other justice topics in Queensland.		– Offences – Courts – Corrections
Department of Communities, Child Safety and Disability	Produces an annual report which provides Queensland youth justice services data.		– Annual report – Child protection statistics

Services				
Crime and Corruption Commission (CCC)	Produces an annual report of yearly activities and conducts research projects in policing, illicit drugs, paedophilia, capacity development and crime prevention.		– Annual report – Publications and research ○ Abuse ○ Crime and law enforcement ○ Police ethics ○ Property crime ○ Tasers	○ Drugs ○ Fraud ○ Indigenous ○ Police Misconduct
Key Centre for Ethics, Law, Justice and Governance, Griffith University	Interdisciplinary research centre that has developed a model which simulates the juvenile justice system in Queensland.	– Justice Modelling @ Griffith – Griffith Youth Forensic Service – Prevention and developmental pathways		

Appendix E: Data collection method considerations

Figure 6 in the CJEF indicated that a number of factors need to be considered when selecting an appropriate data collection method for your evaluation. Table E.1 provides a summary of the interaction between each of these key considerations and the various data collection methods. A '✓' indicates the presence, in general, of a particular characteristic of that data collection method. For example, both written and online surveys are able to sample a large number of respondents. Please note that Table E.1 is only intended as a guide – there will always be exceptions to the rule.

Table E.1. Data collection method considerations

Data collection methods	Large sample size	Efficient	More expensive	Standardised data	Objective	Flexible	Privacy	Confidential	Allows clarification	Rich data	Evaluator ease	Respondent ease	Utilises existing data	Data completeness
Surveys/questionnaires														
Written survey	✓	✓		✓	✓	✓	✓	✓			✓	✓		
Online survey	✓	✓		✓	✓	✓	✓	✓			✓	✓		✓
Interviews														
One-on-one interview			✓	✓	✓	✓	✓		✓					✓
Telephone interview	✓	✓	✓	✓	✓	✓	✓		✓			✓		✓
Focus group		✓				✓	✓		✓	✓				✓
Group assessment		✓		✓	✓	✓	✓		✓	✓				✓
Community forum	✓	✓				✓	✓		✓	✓				✓
Submission/testimonial	✓	✓				✓	✓			✓	✓			✓
Document reviews														
Expert or peer review	na			na			na	na		✓	✓	na	✓	na
Literature review	na			na			na	na		✓	✓	na	✓	na
Program/policy documents review	na			na				na		✓	✓	na	✓	na
Official records review	✓	✓		✓	✓			✓			✓	na	✓	
Log/journal/diary review	na			na				na		✓	✓	na	✓	na
Observation			✓			✓	✓			✓		✓		✓

Data collection methods	Large sample size	Efficient	More expensive	Standardised data	Objective	Flexible	Privacy	Confidential	Allows clarification	Rich data	Evaluator ease	Respondent ease	Utilises existing data	Data completeness
Performance measures														
Test of knowledge	✓	✓		✓	✓	✓	✓	✓			✓	✓		
Simulated problem or situation			✓	✓		✓	✓		✓	✓				✓
Activity sampling			✓	✓	✓	✓	✓		✓	✓		na		✓
Rating	✓	✓		✓		✓	✓	✓			✓	✓		

Appendix F: Ethical issues in criminal justice evaluations

Evaluations that involve offenders as participants or subjects, or investigate illegal activities, need to consider a number of unique ethical issues. These include limits to confidentiality in certain circumstances; coercion in dependent participant-evaluator relationships; ensuring informed consent is gained; providing adequate mechanisms for raising concerns or complaints; and ensuring the need for privacy.

Limits to confidentiality

Informing potential participants of any limits to confidentiality is a particularly important issue. There are a number of circumstances in which confidentiality cannot be guaranteed. Mandatory reporting requirements mean that suspected child abuse must be reported to authorities. Similarly, details of non-adjudicated criminal offences and potential harm to self or others may be required to be reported to authorities. Other situations may also be relevant depending on your evaluation. Evaluators should clearly inform potential participants about issues of disclosure in plain language. Some examples of statements explaining limits to confidentiality are:

- *You should not disclose specific information about illegal behaviours that you have not been charged with or have not been dealt with by a court.*
- *Researchers cannot guarantee absolute confidentiality in relation to illegal behaviours of which they are made aware.*

Dependent relationships

In criminal justice evaluations, the possibility exists that participants may be recruited in situations where they are in some way dependent upon the person doing the recruiting, or where there is an unequal relationship (e.g., police officer/offender, counsellor/client). The National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Research Involving Humans indicates that the consent of a person to participate in research must not be subject to any coercion. Determine how potential problems arising from unequal or dependent relationships between recruiters and participants will be handled. For example, if the research involves participants in a dependent relationship with the researcher, then consent must be witnessed by an independent person.

Incentives for participants

The appropriateness of offering incentives to inmates or offenders as inducement to participate in an evaluation is also contentious given that inmates are in vulnerable and dependent relationships within the correctional system. Inadvertent coercion of the inmate to participate is possible. In particular, the inmate may feel a tacit obligation to answer all questions, or that the incentive will be withdrawn if they do not fully co-operate. Additionally, the individual may expect that a financial inducement of comparable value will be offered by all agencies whenever they are asked to participate in research, and this jeopardises evaluations undertaken by some agencies where such incentives cannot be accommodated within their budget.

It may be appropriate to offer incentives in certain circumstances, for example, where a financial payment is considered a reimbursement for an inmate's lost wages from prison employment.

Ensuring informed consent

Offenders and prisoners often have lower than average numeracy and literacy levels, so participant information and consent forms must be written in plain, concise language that participants will understand. Given that offenders are prisoners may also be unlikely to disclose a lack of understanding, evaluators may consider reading participant information aloud, or utilising other means to ensure that the information has been understood.

Raising concerns or complaints

A statement must be included that details how participants can raise concerns or complaints about the conduct of the project, and must provide contact details for the relevant ethics committees. Prisoners should be advised to contact the official prison visitor in the first instance if they have any queries, concerns or complaints about the conduct of the evaluation.

Privacy issues

Individuals must provide consent for their personal data (i.e., offender files, criminal histories) to be accessed. If consent is not gained, data must be modified and provided in de-identified form. De-identified data means information that does not reasonably identify the individual, or from which the person's identity cannot reasonably be ascertained. Information that may identify an individual includes information that is unique in some way or highly specific, for example, name, address or other contact details, or date/place of birth.

Data may be disclosed to an evaluator without information that could identify the individual, but coded so that it may be re-identified if necessary. If the researcher does not have access to the code, then the information collected and subsequently used by the researcher is **de-identified**. If the researcher is given the code, as well as the information, then the information is **potentially identifiable**.

The use of unique identifiers, or potentially identifiable information, is dealt with in privacy legislation. Researchers should ensure that the use of identifiers is done in accordance with any relevant privacy principles that deal with identifiers. Queensland public sector agencies are required to adhere to the privacy regime that is contained in *Information Standard 42* (IS42).[27] If the evaluation utilises data held by an organisation in the Commonwealth public sector, Commonwealth legislation also applies.[28]

Appendix G: Free and open-source software for evaluation

The following is a list of free and open-source resources to support research and evaluation activities. It is not meant to be exhaustive, but represents some of the better supported and most actively developed products which cover a variety of functions.

Name	Web site [and platform]	Comments
Data entry tools		
EpiInfo	http://www.cdc.gov/epiinfo/ [Windows]	*EpiInfo*, from the United States (US) Centre for Disease Control (CDC), provides an integrated package for data entry, analysis and reporting for epidemiological data. The data entry component, *Enter*, enables the rapid creation of formatted data entry screens, and the automatic creation of a *Microsoft Access* database for the storage of the data.
CSPro	http://www.census.gov/ipc/www/cspro/index.html [Windows]	*CSPro* (Census and Survey Processing System) from the US Census Bureau is a public-domain software package for entering, editing, tabulating and mapping census and survey data. Used by organisations such as the World Bank, data is stored in simple text files.
EpiData	http://www.epidata.dk/ [Windows]	*EpiData* is another program for the creation of formatted data entry screens and, like *CSPro*, stores data in text files.
Transcribing tools		
Transcriber	http://trans.sourceforge.net/en/presentation.php [Windows]	A tool for assisting the transcription of recorded material.
Express Scribe	http://www.nch.com.au/scribe/ [Windows]	A tool for assisting the transcription of recorded material.
TAMS Analyzer	http://tamsys.sourceforge.net/ [OSX]	*Text Analysis Mark-up System* (TAMS Analyzer) is an open-source qualitative package for the analysis of textual themes. It can be used for transcribing digital media and conducting discourse analysis in the social and cultural sciences.
Quantitative data analysis		
R	http://www.r-project.org/ [Linux, Windows, OSX, others] Australian mirror sites and additional information: http://cran.ms.unimelb.edu.au/ Search documentation and help archives: http://finzi.psych.upenn.edu/search.html	The *R* language is a free software environment for statistical computing and graphics. *R* is an open-source product, created and supported by volunteers around the world. It has a core of base functions and literally hundreds of add-on packages available for specialist tasks (e.g., epidemiology, meta-analyses, SEM, mixed effects modelling). *R* can import data from a variety of proprietary formats and directly from databases.
	Add-on graphical user interface for *R*: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/ [Cross-platform (Tcl/Tk)]	While *R* in its native form is command-line based, there are a number of projects providing graphical user interfaces to *R*, which are very useful tools when learning *R*. *R Commander* is one of the most developed, and easiest to install, because it is available as an add-on package. It allows access to a wide range of data import, statistical and graphing functions through a familiar menu system, and doesn't hide the underlying code.

	Syntax editor for *R*: http://www.sciviews.org/Tinn-R/ [Windows]	*Tinn-R* is an editor for *R* code, with built-in syntax highlighting and short cuts to run code selections. It can be used alongside the *R Commander*.
EpiInfo	http://www.cdc.gov/epiinfo/ [Windows]	*EpiInfo Analysis* has a range of functions suitable for epidemiological projects.
Gretl	http://gretl.sourceforge.net/ [Linux, OSX, Windows]	*Gretl*, the Gnu Regression, Econometrics and Time-series Library, is a software package for econometric and time series analyses. It can link to R for further functionality.
Sample size and power calculators		
Piface	http://www.cs.uiowa.edu/~rlenth/Power/ [Cross-platform (Java)]	Java applet for power and sample size calculations.
SampleXS	http://www.brixtonhealth.com/samplexs.html [Windows]	*SampleXS* calculates random, systematic or complex samples for cross-sectional surveys.
Sample size calculator	http://www.macorr.com/ss_calculator.htm [Windows]	A simple sample size calculator.
G*Power 3	http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ [Windows, OSX]	*G*Power 3* offers statistical power analyses for many different statistical tests.
Data mining		
R package 'Rattle'	http://rattle.togaware.com/ [Cross-platform]	*Rattle* (the R Analytical Tool To Learn Easily) provides a simple and logical interface for quick and easy data mining. It is being rapidly developed, and is in use by the Australian Taxation Office.
Weka	http://www.cs.waikato.ac.nz/ml/weka/ [Cross-platform (Java)]	*Weka* is a Java-based data mining interface that is supported by the developers' text.
Tanagra	http://eric.univ-lyon2.fr/~ricco/tanagra/ [Windows]	*Tanagra* is free data mining software for academic and research purposes.
Clinical audit tools		
Auditmaker	http://www.auditmaker.org/ [Windows, requires Microsoft Access]	*Auditmaker* is a tailored *Microsoft Access* database for conducting clinical audits from the Australian Centre for Evidence Based Clinical Practice, who also offer training.
Qualitative data analysis		
AnSWR	http://www.cdc.gov/hiv/software/answr.htm [Windows]	*AnSWR* (Analysis Software for Word-based Records) comes from the CDC, and can be used for the qualitative coding of any text-based source material.
Weft-QDA	http://www.pressure.to/qda/ [Windows, OSX]	*Weft QDA* is an easy-to-use graphical user interface package for the analysis of unstructured textual data such as interviews and fieldnotes.
TAMS Analyzer	http://tamsys.sourceforge.net/ [OSX]	*Text Analysis Mark-up System* (TAMS Analyzer) is an open-source qualitative package for the analysis of textual themes. It can be used for transcribing digital media and for conducting discourse analysis in the social and cultural sciences.
Graphing		
EpiInfo	http://www.cdc.gov/epiinfo/	*EpiInfo* has a variety of graphing functions.

Ggobi	http://www.ggobi.org/	*GGobi* is an open-source visualization program for exploring high-dimensional data. It enables animation such as rotations of 3 dimensional scatterplots.
Tee Tree Office	http://www.steema.com/products/teetree/office/overview.html [Windows]	A freeware tool for drawing flowchart diagrams.
Tee Chart Office	http://www.steema.com/products/teechart/office/overview.html [Windows]	A freeware charting tool.
Epigram	http://www.brixtonhealth.com/ [Windows]	A general purpose diagram creation tool that is easy to use.
Project management tools		
Ganttproject	http://ganttproject.biz/ [Cross-platform (Java)]	A Java-based *gantt* charting tool.
Freemind	http://freemind.sourceforge.net/wiki/index.php/Main_Page [Cross-platform (Java)]	A Java-based mind-mapping software that will export to *Ganttproject*.
ToDoList	http://www.abstractspoon.com/ [Windows]	A 'to do' list manager that will export to *Ganttproject*.
Planner	http://winplanner.sourceforge.net/ [Windows] http://www.simpleprojectmanagement.com/planner/home.html [Linux]	Another project planning and *gantt* charting application.
Other useful tools		
OpenOffice.org	http://openoffice.org [Linux, Windows, OSX, others]	An open-source office suite (i.e., word processor, spreadsheet, presentation and database applications) that is compatible with *Open Document* and *Microsoft Office* formats.
Bibus	http://bibus-biblio.sourceforge.net/wiki/index.php/Main_Page [Cross-platform (Python)]	A bibliographic database that will integrate with *OpenOffice.org* and *Microsoft Office*, and import references from *Endnote*.
Online lists of free software		**Online statistics textbooks**
– http://freestatistics.altervista.org/en/stat.php – http://statpages.org/javasta2.html – http://www.statsci.org/free.html – http://www.hmdc.harvard.edu/micah_altman/socsci.shtml – http://gsociology.icaap.org/methods/		– http://www.statsoft.com/textbook/stathome.html

Source: M. Bibo, Queensland Health, personal communication, April 20 2007.

Appendix H: Evaluation report structure

A suggested evaluation report structure includes:

Contents

Executive summary

Introduction

- Purpose of evaluation
- Location of the evaluation
- Description of what was evaluated
- Strategy/service delivery outputs
- Scope/scale of evaluation
- Methodology/approach, including data gathering and analysis tools
- Constraints on the study, including time, cost, expertise, credibility, political and social environment

Findings and conclusions

- Inclusion of information and data collected
- Documentation and analysis of the information and stakeholder views
- Documentation and discussion of findings and conclusions, including: performance of each activity in terms of outputs; performance against outcomes; cost and efficiency; and impact on target group.

Recommendations

Appendices

- Detailed documentation of data collection and analysis procedures
- List of references
- List of departments consulted during the study
- List of evaluation steering committee members

Appendix I: Line, bar and pie graphs

Graphs help to summarise and illustrate information concisely and clearly. In general, line graphs display larger quantities of information, and more detail, than bar or pie charts. Table I.1 presents advice on when to use the various graphical forms, and some considerations to keep in mind during their development.

Table I.1. Line and bar graphs: Their appropriateness and some considerations

	Appropriate for	Considerations
Line graphs	– Representing a continuous variable (e.g., time) on the horizontal axis. – Displaying more than one relationship in the same diagram (e.g., interactions, relationship between three variables).	– Unless the lines are well separated, graphs with more than four or five lines tend to become confusing – Use different line styles, colours or plotting symbols to distinguish lines in a graph with more than one – Use line styles, colours or plotting symbols consistently in any set of line graphs – Consider using the same scale when comparisons are to be made across graphs
Bar charts	– Representing a discrete variable or categorisation (e.g., treatment group) on the horizontal axis.	– Decide upon an order to the bars (e.g., by time, descending order). – In a series of charts, keep the bar order and shading consistent. – To highlight certain comparisons, cluster or group the bars according to the categories they represent. – 'Stacking' bars enables charts to display more complex information.
Pie charts	– Displaying information about the size of classes or groups (or percentages) in proportion to a whole (i.e., 100% or 360 degrees).	– Each section of the pie corresponds to a category of the variable represented (e.g., each age group), and the size of the section is proportional to the percentage of the corresponding category. – Use percentages, proportions, or totals to label the sections of the chart. When comparisons are to be made across graphs, use the same labelling scheme. – In a series of charts, keep the section order and shading consistent. – Consider combining data groups when each amount to less than 5% of the pie.

Appendix J: Internal or external evaluator?

The flowchart in Figure J.1 is designed to help you to decide whether an internal or external consultant should conduct the evaluation.

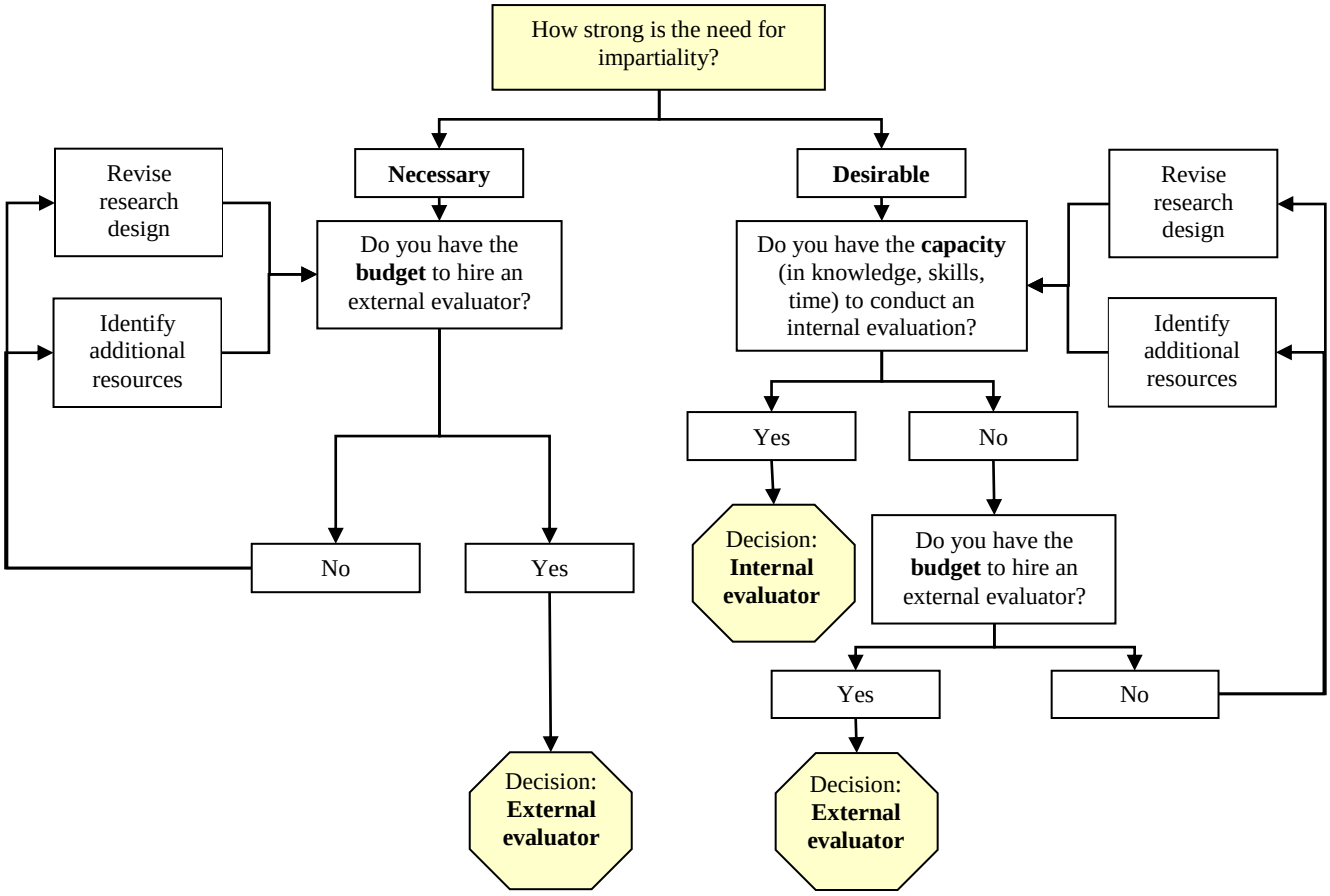

Figure J.1. Process for deciding between an internal and external evaluator

Appendix K: External consultant contractual agreement

A comprehensive contractual agreement in place between those commissioning the evaluation and the external consultants would specify the following:[29]

Basic considerations	Object of the evaluation; purpose of the evaluation; client; stakeholders and interested parties; authorised evaluator(s); guiding values and criteria; standards for judging the evaluation; contractual questions.
Information	Required information; data collection procedures; data collection instruments and protocols; information sources; participant selection; provisions to obtain needed permissions to collect data; follow-up procedures to assure adequate information; provisions for assuring the quality of obtained information; provisions to store, maintain security and dispose of collected information.
Analysis	Procedures for analysing quantitative and qualitative data.
Reports	Deliverables and due dates; interim report formats, contents, lengths, audiences, and methods of delivery; final report format, contents, length, audiences, and methods of delivery; restrictions/permissions to publish information from or based on the evaluation
Reporting safeguards	Anonymity/confidentiality/privileged communication; pre-release review of reports; editorial authority; final authority to release reports; ownership of materials and intellectual properties; any subsequent use of evaluation materials.
Protocol	Contact persons; rules for contacting program personnel; communication channels and assistance.
Evaluation management	Time line for evaluation work of both clients and evaluators; assignment of evaluation responsibilities.
Client responsibilities	Access to information; services; personnel; information; facilities; equipment; materials; transportation assistance; work space.
Evaluation budget	Payment amounts and dates; conditions for payment, including delivery of required reports; budget limits/restrictions; agreed-upon indirect/overhead rates; contracts for budgetary matters.
Review and control of the evaluation	Contract amendment and cancellation provisions; provisions for periodic review, modification, and renegotiation of the evaluation design as needed; provision for evaluating the evaluation against professional standards of sound evaluation; procedures for dealing with any disputes that may arise.

Appendix L: Guidelines for the ethical conduct of evaluations

The following is adapted from the Guidelines for the Ethical Conduct of Evaluations, developed by the Australian Evaluation Society. It outlines certain procedures that you can adopt whilst planning, conducting and reporting on your evaluation which help you adhere to ethical principles.

Planning for an evaluation

Look for potential risks or harms

Anticipate and discuss potential risks or harms to the clients, program staff, or other groups that are involved at the beginning of an evaluation.

Assess the potential effects and implications of the evaluation, both positive and negative, on all involved stakeholders.

Practice within your competence

Evaluators should possess the knowledge, abilities, skills and experience that are needed to undertake the evaluation, and represent their competence fairly.

Compete honourably

When evaluators apply to tender for an evaluation, they should conduct themselves in a professional and honourable manner.

Deal openly

Those tendering an evaluation should deal with all proposals openly and fairly, and respect ownership of materials, intellectual property and commercial confidence.

Disclose potential conflicts of interest

When evaluators apply to tender for an evaluation, they should disclose any of their roles or relationships that may create potential conflict of interest. Mention should also be made to this conflict in any evaluation documents, including the final report.

Conducting an evaluation

Obtain informed consent

All those who provide information to the evaluation should be advised about what information is sought, how the information will be recorded and used, and the likely risks and benefits arising from their participation in the evaluation. Following this, the informed consent should be sought from all, preferably in writing. In the case of minors or other dependents, informed consent should also be sought from parents or guardians.

Be rigorous

The design, data collection methods and data analyses of the evaluation should be rigorous, and in line with the objectives of the evaluation.

Maintain confidentiality

All information collected during the evaluation should be stored and disposed of confidentially. The results or findings should be treated confidentially until released.

Reporting the results of an evaluation

Report clearly and simply

All evaluation results should be presented clearly and simply to allow stakeholders to easily understand the evaluation process and results. If tailoring reports or other communications to a given stakeholder, each should include all important results.

Report fairly and comprehensively

Evaluation reports should be direct, comprehensive and honest in the disclosure of findings and the limitations of the evaluation. Reports should interpret and present evidence and conclusions in a fair manner, and include sufficient details of their methodology and findings to substantiate the conclusions.

Identify sources and make acknowledgements

The source of all comments, contributions and conclusions should be easily identified and acknowledged, unless anonymity is requested.

Appendix M: Risk management

Table M.1 provides some examples of potential risks to evaluation success, and appropriate preventive and contingent strategies.

Table M.1. Potential risks to evaluation success, and preventative and contingent risk management strategies.

Potential risk	Risk management strategies	
	Preventive	Contingent
Agency commitment	Highlight benefits of the evaluation to agencies. Maintain regular contact with agency representatives to ensure their needs are met.	Allow sufficient timeframes and resources to ensure that agencies have sufficient time to respond to requests, and modify as required.
Timing/scheduling of activities	Regular monitoring of project plan milestones by evaluators. Allocate sufficient time between project milestones to allow for a small amount of project creep. Communicate with other evaluations/reviews to avoid duplication and to promote complementary outcomes, where possible and appropriate.	Increase frequency or change communication/reporting timelines.
Lack of interest from external consultants/evaluators	Communicate with universities and other reputable research organisations prior to formal tender process.	Meet with prospective consultants individually to outline research requirements.
Lack of reliable data	Collaborate with existing initiatives/evaluations that have required modifications to existing data collections. Clearly communicate data requirements at the earliest possible time. Where appropriate, seek funding for the development of individual evaluation databases.	Attempt to access similar data from an alternative source. Modify research design.
Evaluation does not impact upon policy development	Include key policy decision-makers on evaluation committee.	Explore the viability of direct links to the policy environment, for example, in CBRC Decisions.
Inability to secure funding for evaluation	Ensure funding proposals accord with established budgetary processes.	Explore leverage options to extend government funding.
Conflict of interest for external evaluators	External evaluators will not determine evaluation objectives.	External evaluators should indicate a perceived conflict of interest where appropriate.

Appendix N: Program Evaluation Standards

The *Program Evaluation Standards*, identified by 16 North American professional associations, are principles which aim to guide the design, conduct and assessment of program evaluations. The standards address the utility, feasibility, propriety, and accuracy of evaluations.

Utility Standards

The utility standards ensure that an evaluation will serve the information needs of stakeholders.

Stakeholder identification — Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

Evaluator credibility — The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

Information scope and selection — Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders

Values identification — The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

Report clarity — Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

Report timeliness and dissemination — Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

Evaluation impact — Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Feasibility Standards

The feasibility standards ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

Practical procedures — The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

Political viability — The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

Cost effectiveness	The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

Propriety Standards

The propriety standards ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

Service orientation	Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.
Formal agreements	Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.
Rights of human subjects	Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
Human interactions	Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.
Complete and fair assessment	The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.
Disclosure of findings	The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.
Conflict of interest	Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.
Fiscal responsibility	The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Accuracy Standards

The accuracy standards ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

Program documentation	The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.
Context analysis	The context in which the program exists should be examined in enough

	detail, so that its likely influences on the program can be identified.
Described purposes and procedures	The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
Defensible information sources	The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
Valid information	The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.
Reliable information	The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.
Systematic information	The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.
Analysis of quantitative information	Quantitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.
Analysis of qualitative information	Qualitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.
Justified conclusions	The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
Impartial reporting	Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.
Metaevaluation	The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Source: The Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards*. Thousand Oaks, CA: Sage Publications, Inc.

Appendix O: Evaluation plan checklist

Upon completion of evaluation planning, the plan should be compared to the following evaluation plan checklist. The checklist examines critical aspects of the plan and the relationship among their parts.

Focusing the evaluation

- Has an evaluation committee been formed? ❑
- Does the evaluation plan contain a logic model? ❑
- Are all evaluation inputs, outputs and outcomes listed and measurable? ❑
- Are links between the logic model, research form and research questions explicit? ❑
- Are variables identified for each evaluation question? ❑
- Will the selected research methods answer the research questions? ❑
- Are data analyses appropriately matched to the research questions? ❑

Variables and data

- Is the chosen sample unbiased and of optimal size? ❑
- Are appropriate and feasible data collection methods detailed for each variable? ❑
- Is data granularity and frequency appropriate? ❑
- Are there appropriate methods/measures for each variable? ❑
- Are data sources valid and feasible? ❑

Data management

- Have data sharing protocols been established? ❑
- Has ethics approval been gained? ❑
- Have plans for storing data been made? ❑

Reporting strategy

- Have reporting timelines been documented? ❑
- Has the report format and structure been determined? ❑
- Are plans in place for displaying data information results? ❑

Managing the evaluation

- Is a contractual agreement in place to manage the tender process? ❑
- Have risks been identified and strategies developed for their management? ❑
- Has the evaluation timeframe, budget and project plan been finalised? ❑

Appendix P: Evaluation plan template

Evaluation Plans need to clearly focus on the purpose of the evaluation, the issues that are to be addressed, and fundamental evaluation considerations. Evaluation Plans are not intended to be static documents and should be reviewed regularly throughout the evaluation of a policy, program or initiative.

A recommended Evaluation Plan template for CBRC submissions is provided below. This template is designed as a tool to assist agencies in considering evaluation issues during the policy development phase. Agencies need not use this particular template, particularly where established project management frameworks are already in place. However, any Evaluation Plan presented to CBRC for consideration should demonstrate that considerable thought has been given to the evaluation, and would also need to address the key issues outlined in this template.

Purpose and rationale

- Briefly describe the overarching purpose and objective(s) of the evaluation, providing sufficient rationale for these and why the evaluation is required (e.g., Ministerial-level request, Cabinet or CBRC decision, intergovernmental agreements, policy commitments, evaluation unit request).

Governance

- Detail what governance structures are in place for the evaluation. Identify the lead agency for, and other agencies involved in delivering, the evaluation. An organisational chart or matrix may be appropriate. This section should include all organisations that have a role in overseeing the evaluation including steering committees, Cabinet, or traditional departmental structures. Detail the lines of accountability and responsibility.
- Provide a rationale for selecting an internal or external evaluator.

Proposed evaluation model

- Specify the proposed evaluation model. It is advisable to identify the evaluation type; key research questions; methodology; what information is needed; proposed data sources or collection methods; sampling considerations; and data sharing and storage arrangements.

Evaluation timing

- Specify the dates for completing key tasks, when key milestones or targets will be reached, and a date that will indicate when the evaluation must be completed.

Reporting arrangements

- Outline the evaluation-specific reporting arrangements. Identify the milestones that will be reported against, the form of the reports, when the reports are due, and the audience for the report (e.g., Premier, Minister, Cabinet, steering committee, public).

Ethical considerations

- Detail relevant ethical considerations which may arise during the planning, conducting and reporting of the evaluation. Also, identify how approval for the evaluation will be sought from a research ethics committee.

Resource considerations

- Make a preliminary determination as to whether the evaluation will be funded internally or additional funding will be required. Detail any relevant Cabinet or CBRC decisions, Ministerial approvals, or departmental decisions that may inform this decision.
- Provide estimated budget and expenditure figures, and make a statement in respect of the cost-effectiveness of the evaluation. Undertake contingency planning and outline the strategies that are in place to ensure that the evaluation stays within budget.

Risk management

- Identify the key risks to achieving the evaluation and deliverables, the likelihood of these occurring, the potential impact if they do occur, and possible strategies to mitigate these risks.

Stakeholder consultation

- List the key stakeholders who will participate in, or be affected by, the evaluation. Identify whether there is agreement among stakeholders regarding the evaluation purpose and model, and whether disagreement is problematic.
- Clarify what, if any, consultation is planned with stakeholders; how it will be conducted, by whom and when; the purpose of consultation and what is hoped to be achieved through consultation; and how any disagreement between stakeholders, or between Government and stakeholders, will be managed.

Endnotes

[1] Australian Evaluation Society. (2006). *Guidelines for the ethical conduct of evaluations.* Available at: http://www.aes.asn.au/about/guidelines_for_the_ethical_conduct_of_evaluations.pdf.

[2] Australian Evaluation Society. (2006). *Guidelines for the ethical conduct of evaluations.*

[3] Model is adapted from English, B., Straton, R.G., & Cummings, R. (2002). *Principles for evaluating community crime prevention projects.* Attorney-General's Department, Commonwealth Government. Available at: http://www.ag.gov.au/agd/www/Ncphome.nsf/Page/9CB89639A8BF1D1BCA256C2900198AA0?OpenDocument.

[4] Pawson, R., & Tilley, N. (1994). What works in evaluation research. *British Journal of Criminology, 34,* 291-306.

[5] National Crime Prevention. (1999). *Pathways to prevention: Developmental and early intervention approaches to crime in Australia*. National Crime Prevention, Attorney-General's Department: Canberra. Available at: http://www.ag.gov.au/agd/www/Ncphome.nsf/Page/B78FEDFB9A1D980ACA256B14001A096E?OpenDocument.

[6] Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2003). *Evaluation: A systematic approach* (7th Ed.). Sage Publications: Thousand Oaks, California.

[7] National Crime Prevention. (1999). *Pathways to prevention: Developmental and early intervention approaches to crime in Australia*. National Crime Prevention, Attorney-General's Department: Canberra. Available at: http://www.ag.gov.au/agd/www/Ncphome.nsf/Page/B78FEDFB9A1D980ACA256B14001A096E?OpenDocument.

[8] Queensland Treasury. (2006). *Cost-benefit analysis guidelines: Achieving value for money in public infrastructure and service delivery*. Available at: https://www.google.com.au/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAAahUKEwjImdP4-4bGAhXk2qYKHeLmAEg&url=https%3A%2F%2Fwww.treasury.qld.gov.au%2Fpublications-resources%2Fprojects-queensland%2Fpolicy-framework%2Fproject-assurance-framework%2Fpaf-cost-benefit-analysis.pdf&ei=nyR5VYjOJuS1mwXizYPABA&usg=AFQjCNGRjjmRHgQhEiaVZ7Z8-i88mhtaVQ&bvm=bv.95277229,d.dGY

[9] The following units within agencies can provide guidance: *Research and Ethics Advisory Unit* within Queensland Health; *Review and Evaluation Unit*, Ethical Standards Command, within Queensland Police Service; *Strategic Policy* within the Department of Justice and the Attorney-General; *Strategic Policy* within the Department of Communities; and *Strategic Policy and Services* within Queensland Corrective Services.

[10] English, B., Straton, R.G., & Cummings, R. (2002). *Guidelines for evaluating community crime prevention projects.* Attorney-General's Department, Commonwealth Government. Available at: http://www.ag.gov.au/agd/www/Ncphome.nsf/Page/RWP4A7E9FB71531D36ECA256E440000C8AF?OpenDocument.

[11] Onwuegbuzie, A.J., & Leech, N.L. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report, 11* (3), 474-498.

[12] The Division of Instructional Innovation and Assessment. (n.d.). *Coding Qualitative Data.* Quick Tip 7.32, University of Texas. Available at: http://www.utexas.edu/academic/diia/assessment/iar/resources/quicktips/quicktip_7-32.pdf.

[13] The Division of Instructional Innovation and Assessment. (n.d.). *Coding Qualitative Data*; Berkowitz, S. (1997). Analyzing qualitative data. In J. Frechtling & L. Sharp Westat, *User-friendly handbook for mixed method evaluations.* National Science Foundation. Available at: http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/CHAP_4.HTM.

[14] Buston, K. (1997). NUD*IST in action: Its use and its usefulness in a study of chronic illness in young people. *Sociological Research Online, 2*. Available at: http://www.socresonline.org.uk/socresonline/2/3/6.html.

[15] Barry, C.A. (1998). Choosing qualitative data analysis software: ATLAS/ti and NUD*IST compared. *Sociological Research Online, 3* (3). Available at: http://www.socresonline.org.uk/3/3/4.html; Stanley, L., & Temple, B. (1995). Doing the business: Using qualitative software packages in the analysis of qualitative datasets. In R.G. Burgess (Ed.), *Studies in qualitative methodology: Computing and qualitative research* (169-193). Greenwich, US: JAI Press.

[16] Taylor-Powell, E., & Steele, S. (1996). *Collecting evaluation data: An overview of sources and methods*. Program development and evaluation, Cooperative Extension, University of Wisconsin. Available at: http://learningstore.uwex.edu/pdf/G3658-4.pdf.

[17] Marshall, J. (2005, June 6-7). *Criminal justice data: Limitations and implications for policy research*. Paper presented at the Australian Institute of Criminology Conference, Canberra. Available at: http://www.aic.gov.au/conferences/2005-abs/marshall.pdf.

[18] See, for example, DSS Research. (2006). *Statistical power calculator*. Available at: http://www.dssresearch.com/toolkit/spcalc/power.asp; Kraemer, H. C., & Thiemann, S. (1987). *How many subjects: Statistical power analysis in research*. Newbury Park, CA: Sage Publications.

[19] Australian Institute of Health and Welfare. (2004). *Appendix E: Draft linkage documentation*. In Statistical data linkage in community services data collections. Community Services Ministers' Advisory Council. Available at: http://www.aihw.gov.au/publications/index.cfm/title/10021.

[20] Department for Constitutional Affairs. (2003). *Public sector data sharing: A guide to data sharing protocols*. United Kingdom. Available at: http://www.foi.gov.uk/sharing/toolkit/infosharing.htm.

[21] Commonwealth Government. (2002). *Style manual for authors, editors and printers* (6th ed.). Milton, Queensland: John Wiley and Sons Australia.

[22] See, for example, *The Chicago manual of style* (15th ed.). (2003). Chicago: University of Chicago Press. Electronic guide available at: http://www.chicagomanualofstyle.org; Chapman, M., & Wykes, C. (1996). *Plain figures* (2nd ed.). London: H. M. Stationery Office.

[23] Australian Evaluation Society. (2006). Guidelines for the ethical conduct of evaluations.

[24] For further information, visit http://www.health.qld.gov.au/ethics or contact the Queensland Health Research and Ethics Advisory Unit, Reform and Development Division, at GPO Box 48, Brisbane, Qld, 4001 or Level 18, 147-163 Charlotte Street, Brisbane, Qld, 4000 or phone (07) 3234 0034.

[25] It is advisable that the following units within agencies are your first points of contact: *Research and Ethics Advisory Unit* within Queensland Health; *Review and Evaluation Unit*, Ethical Standards Command, within Queensland Police Service; *Strategic Policy* within the Department of Justice and the Attorney-General; *Strategic Policy* within the Department of Communities; and *Strategic Policy and Services* within Queensland Corrective Services.

[26] Johnson, A.L. (2004). *Engaging Queenslanders: Evaluating community engagement.*

[27] A copy of this standard and its associated guidelines may be accessed at http://www.iie.qld.gov.au.

[28] For Information Privacy Principles (IPPs), refer to *Guidelines Under Section 95 of the Privacy Act 1988* available at http://www.nhmrc.gov.au/publications/synopses/e26syn.htm; for National Privacy Principles (NPPs), refer to *Guidelines Approved under Section 95A of the Privacy Act 1988* available at http://www.nhmrc.gov.au/publications/synopses/e43syn.htm.

[29] Stufflebeam, D. (1999). *Evaluation contracts checklist.* The Evaluation Centre, Western Michigan University. Available at: http://www.wmich.edu/evalctr/checklists/contracts.pdf.