



Criminal justice evaluation and performance monitoring

What is evaluation?

Cite as: Boxall H, Yates S & Bartels L (2025). *Criminal justice evaluation and performance monitoring – What is evaluation?* Australian National University.

This document was prepared by POLIS: Centre for Social Policy Research, Australian National University (ANU), on behalf of the ACT Justice and Community Safety Directorate (JACS), for the *ACT Criminal Justice Data Capacity Project*. It should be considered in combination with the other resources prepared as part of this project, available at www.act.gov.au.

We gratefully acknowledge the assistance of key staff in JACS, including ACT Corrective Services, the First Nations Justice Branch, Restorative Justice Unit and Justice Reform Branch; ACT Health; and Yeddung Mura. We also appreciate the advice of Associate Professors Sean Colishaw, Monash University, and Elfie Shiosaki, ANU, in preparing the resources for this project.

Contents

Introduction.....	1
What does this resource cover?	1
What is evaluation?.....	1
Why do evaluation?	2
Evaluation design	4
Measuring change – what is ‘good’ and ‘bad’ evidence?.....	4
Realist evaluations.....	5
What is a control group?.....	5
Experimental designs.....	6
Natural or quasi-experiments	7
Non-experimental designs.....	7
Evaluation research methods.....	7
Administrative data.....	8
Case file analysis	10
Client surveys.....	11
Interviews.....	11
Focus groups	12
Observations	12
Aboriginal and Torres Strait Islander Indigenous evaluation	13
Indigenous worldviews.....	13
Resources and frameworks for Aboriginal and Torres Strait Islander evaluation.....	14
The AIATSIS Code of Ethics	14
Indigenous data sovereignty.....	16
Best practice	17

Introduction

This resource is designed to help service providers better understand the role of evaluation, how to contribute to meaningful data collection, and how evaluation findings can be used to improve services and outcomes for clients.

What does this resource cover?

- **The purpose of evaluation** – what evaluation is and why it is important.
- **Types of evaluation** – process evaluations, outcome evaluations, and economic assessments.
- **Evaluation designs** – experimental, quasi-experimental, and non-experimental approaches.
- **Data collection methods** – client surveys, administrative data, case file analysis, interviews, and focus groups.
- **Indigenous evaluation principles** – best practices for culturally appropriate research.

What is evaluation?

Evaluation is a way of assessing a service, intervention or program (we will use the term 'program' to stand in for all these). There are a number of different types of evaluation (see Figure 1). The most common are **process** and **outcome** evaluations.

As we discuss later, outcome evaluations often use experimental or quasi-experimental designs to assess outcomes.

Figure 1: Types of evaluation



These guidelines have been developed to support programs and services funded by ACT Justice and Community Safety Directorate to collect information for the most commonly required purposes: **process evaluations, outcome evaluations, and performance monitoring activities**. Performance monitoring means the day-to-day measurement of performance against contractual obligations, which is generally governed by the terms of the funding contract and is therefore out of the scope of the present guidelines.

Why do evaluation?

Programs funded by governments are increasingly required to collect information and data that can be used for evaluation activities. But quite apart from contractual requirements, evaluation and performance monitoring can also have important benefits for your program and the clients you are engaging with.

1. Understand 'what works' (for whom and under what circumstances)

Evaluation is important to help identify what is working well in a program, as well as what can be changed to improve delivery and outcomes. To do this, you need to:

- collect information about your clients' characteristics (Guideline 3)

- collect information about the activities and outputs that are delivered as part of your program (Guideline 4).
- collect feedback from clients about the program (Guideline 5).

You also need to be able to link this information with the range of outcomes your program is working towards (Guideline 6). Together, this information can help understand ‘what works’, for whom and under what circumstances.

2. Understand and explain why a program is not achieving its intended outcomes

Evaluation can also explain *why* a program may not be achieving its intended outcomes. When a program is less effective than expected, it is usually because of one or more of the following:

- **Implementation failure:** The program was not implemented or delivered as intended, so participants did not ‘receive’ the intervention. This could be because key activities were not implemented due to a lack of resources, like funding, staffing or program guidelines (see Guideline 1), clients did not participate for as long as they needed to (e.g., there was a high drop-out rate), or staff were not trained properly to deliver the intervention.
- **Theory failure:** The program was not delivering the ‘right’ activities and outputs to bring about change for clients (see Guideline 4). This could be because the program was not evidence-based, was targeted at the wrong client group (see Guideline 3) or was not adapted to suit the context in which it was being delivered.
- **Measurement failure:** The evaluation itself was ‘wrong’ and did not accurately measure the outcomes of interest. This might be because evaluators have used unreliable or invalid measures or small samples that limited the types of analysis that could be undertaken, or picked inappropriate intervention and control groups, or did not follow people up for long enough.
- **Environmental failure:** Environmental factors that are often outside the service provider’s control can also affect the effectiveness of a program. This can include the weather for programs that rely on being outside, a Sorry Business event occurring mid-program or even planning a program over periods that include public holidays or school holidays, which may impact participants with school-age children.

These guidelines are intended to minimise the risk that negative evaluation findings are due to measurement failure (Guidelines 3-6).

3. Improve service delivery and improve outcomes for clients

Evaluation activities are also useful for improving service delivery, so using these guidelines may also help you to explain *why* your program may not be achieving its aims (i.e., theory failure or implementation failure).

The information you collect for evaluation can also help in service delivery. In particular, it can help to identify the support needs of clients referred to your program (see Guideline 3) and the interventions that most often resulted in positive outcomes for your clients (Guideline 4 and Guideline 6).

Evaluation design

In this section, we provide some information about different types of evaluation design. Planning work is done before the collection of information starts and you may want to engage the services of an evaluator, if your organisation does not have these skills. This section explains some of the decisions to be made in planning your evaluation.

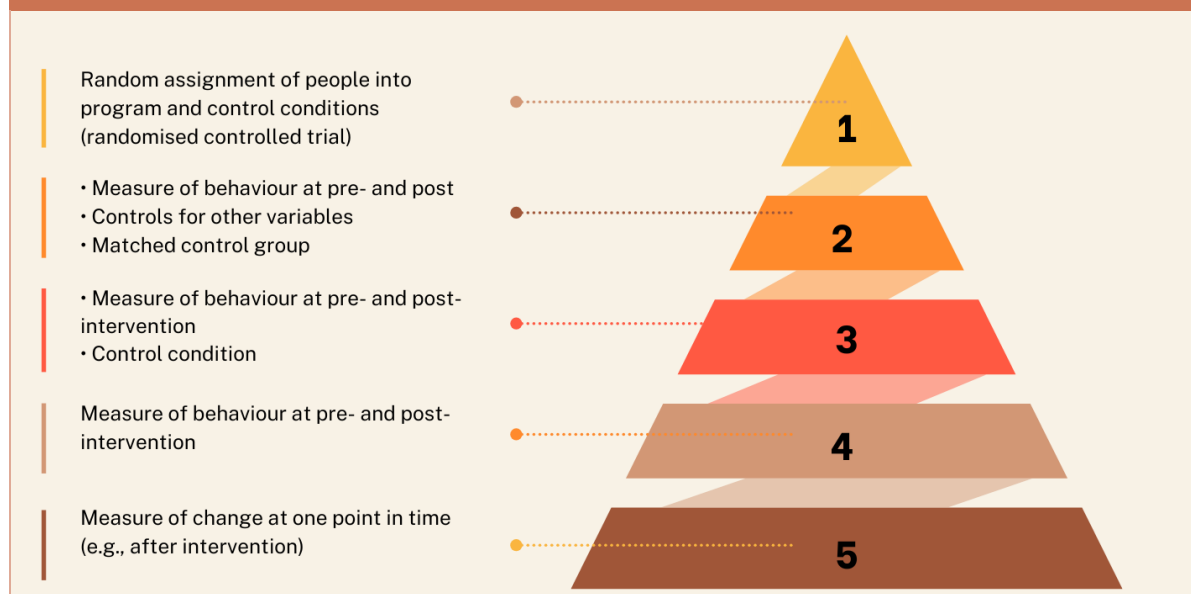
Measuring change – what is ‘good’ and ‘bad’ evidence?

When thinking about evaluation, we need to be mindful that some types of evidence are considered ‘better’ (more rigorous) than others. The Maryland Scientific Methods Scale (see Figure 2) was designed to assess the strength of scientific evidence and methodological quality of outcome evaluations.

It classifies evaluations into one of 5 levels – 5 being the lowest quality methodology and 1 the highest. As a standard principle, evaluations should aim to meet a minimum standard of 3 to be considered robust. Importantly, evaluations that meet the criteria for 1, 2 or 3 include a control or comparison group.

What is considered good or bad evidence depends on the audience. However, as a result of colonisation, there is a tendency to favour methodologies that use Western tools and metrics. This does not mean that this information is actually better than other ways of knowing and doing, especially when working with Aboriginal and Torres Strait Islander clients. We encourage you to look at the section on [Aboriginal and Torres Strait Islander evaluation](#) (in this Guideline) and Guideline 7 on case studies for alternative ways of thinking about evaluation and evidence.

Figure 2: The Scientific Methods Scale¹



¹ D Farrington et al. (2013). ‘The Maryland Scientific Methods Scale’ in *Evidence-based crime prevention*. Routledge: p.13.

Realist evaluations

While not placed at the apex of the Scientific Methods Scale, realist evaluation is important for evaluating criminal justice programs in the ACT. This type of evaluation focuses on understanding how causal mechanisms are contextual – that is, shaped by the particular contexts of the program.²

Realist evaluation helps explain:

- What mechanisms are at play? (i.e., the processes that drive change).
- Which contexts influence these mechanisms? (i.e., the setting, population, and conditions affecting success).
- What outcomes are produced? (i.e., the results that occur for different groups).

For example, men and women might respond differently to the information and resources provided in a prison diversion program. Realistic evaluation will help explain the reasons for such outcomes.

Usually, realist evaluations collect both quantitative and qualitative data, with quantitative data used to measure context and outcomes, while qualitative data explain the mechanisms at work to create the outcomes in the relevant context (see below on research methods). Ideally, realist evaluation should produce ‘context-mechanism-outcome’ statements, explaining which context and mechanisms operated in which way, for which actors, to produce which outcomes.

Realist evaluations do not use comparison groups, so they are generally only rated as Level 3 or 4 on the Scientific Methods Scale. However, there is growing recognition that it is not enough to simply know whether a program works (for example, through an RCT). We also need to understand *how* and *why* it works. These guidelines are therefore informed by realist evaluation principles, as they are intended to support understanding of the mechanisms that underpin programs.

What is a control group?

A control group (or comparison group) is made up of people (or sites, organisations, etc.) who did not receive the intervention but are otherwise similar to those who did. The purpose of a control group is to estimate what would have happened if the program had not been implemented—often referred to as the ‘business as usual’ scenario. **Business as usual** may mean receiving no intervention at all or an alternative program.

By comparing the outcomes of the control group with those of the intervention group, evaluators can determine whether the program made a difference.

The way a control group is selected depends on the evaluation design. In randomised controlled trials ([RCTs](#)), the control and intervention groups are selected randomly. Meanwhile, in [quasi-experimental](#) and [non-experimental](#) designs, control groups are identified in other ways, such as:

² E Owusu-Addo et al. (2019). Cash transfers and the social determinants of health: Towards an initial realist program theory. *Evaluation*, 25(2): p.224.

- People who were referred to the program but chose not to engage or exited before completing the program;³
- People who would have been referred but were not eligible (e.g., the program was not available in their area at the time⁴); or
- People who could have been referred but were not, and who are matched to the intervention group based on key characteristics (e.g., age, gender, prior contact with the criminal justice system⁵).

In an RCT, your service is not responsible for identifying or selecting which clients will be allocated to the control group, as an RCT evaluation is likely to be run by an external party who would do this task. The data you are collecting will be of value to the evaluator – for example:

- client characteristics (see Guideline 3) – helps evaluators identify a suitable control group.
- baseline data (see Guideline 6) – captures participants' starting points, making it possible to measure changes over time.

By collecting this information, your service will help develop rigorous and credible evaluations that assess the impact of your program.

Experimental designs

RCTs are often called the '**gold standard**' in evaluation because they *randomly* assign people to the intervention group, which receives the program and the control group, which does not. They then compare outcomes between those two groups to identify whether the program made a difference.

Randomisation helps solve the problem of comparing very different groups of people. Without randomisation, people who participate in a program may have characteristics that influence their outcomes, making it difficult to tell whether the program itself caused the changes. For example, if people who choose to take part in a restorative justice program already have lower reoffending risks, simply comparing them to those who do not participate might give misleading results. However, if people are randomly assigned to the program, these differences are spread evenly across groups, ensuring a fair comparison.

There are some debates about whether RCTs are truly the 'gold standard', because randomisation may not in fact control for differences that could explain any observed differences in outcomes. As such, some researchers suggest that using matched control groups and quasi-experimental designs produce more robust evidence.⁶

³ D Weatherburn et al. (2020). *The Long-Term Effect of the NSW Drug Court on Recidivism*. New South Wales Bureau of Crime Statistics and Research.

⁴ See: H Boxall et al (2020). *Responding to Adolescent Family Violence: Findings from an Impact Evaluation*. Australian Institute of Criminology.

⁵ See: S Lawler et al. (2025). *Restorative justice conferencing for domestic and family violence and sexual violence: Evaluation of Phase Three of the ACT Restorative Justice Scheme*. Australian Institute of Criminology.

⁶ Grossman, J & Mackenzie, F (2005). 'The Randomized Controlled Trial: Gold Standard, or Merely Standard?' *Perspectives in Biology and Medicine* 48(4):p. 516; Parra, J & Edwards Jr, D (2023). 'Challenging the Gold Standard Consensus: Randomised Controlled Trials (RCTs) and Their Pitfalls in Evidence-Based Education' *Critical Studies in Education*, 25(5):513-30..

RCTs also raise ethical challenges – should people in need be denied access to a potentially helpful program just for the sake of evaluation? In response to this concern, one option is to use ‘wait-list’ controls. People in the wait-list control group, are merely delayed (not denied) access to the program. This approach allows evaluators to compare outcomes, while ensuring everyone eventually benefits from the program. RCTs sit at Level 1 on the Scientific Methods Scale (Figure 2).

Natural or quasi-experiments

Often, randomisation is often not possible (or ethical), so evaluators try to identify a control group that is as similar as possible to the intervention group and examine outcomes for that group as well. For example, the evaluation of outcomes for the NSW Drug Court compared people who were accepted into the program with a similar group of people who were referred to the program but not accepted, while controlling for factors such as age, sex, principal offence and number of prior convictions.⁷

This design is known as a quasi-experiment or natural experiment, because researchers are using naturally-occurring circumstances in a field setting, rather than deliberately manipulating the experimental groups.⁸ Quasi-experimental designs are typically Level 2 or 3 on the Scientific Methods Scale (Figure 2).

Non-experimental designs

Finally, non-experimental designs do not involve the use of a control or comparison group and so only meet Level 4 or 5 on the Scientific Methods Scale. Common non-experimental designs include case studies (see Guideline 7), pre- and post-testing and realist evaluation.

Pre-and post- testing involves collecting data at least two times during a client’s engagement in the service; typically, this includes before their involvement in the service or during the initial period of engagement and on program exit. The Client Outcomes Survey (see Guideline 6) is designed to be used at these timepoints with clients referred to your service.

Evaluation research methods

Evaluations can use various research methods that involve collecting quantitative and/or qualitative data. Quantitative data involves numbers (e.g., how many people in your program were 18 years old?), while qualitative data is usually text-based (e.g., ‘what did participants like best about this program?’).

Importantly, one research method can create both quantitative and qualitative data. For example, a survey of clients may include both closed-ended questions, yielding quantitative data (e.g., how old are you right now?) and open-ended questions, where clients can write their response (e.g., what did you like about the program?). The table below sets out the types of evaluation questions that can be answered with different research methods

⁷ D Weatherburn et al. (2020). *The Long-Term Effect of the NSW Drug Court on Recidivism*, New South Wales Bureau of Crime Statistics and Research.

⁸ Simpson, R (2024). ‘Experimental Methods in Criminology’ in *Oxford Research Encyclopedia of Criminology and Criminal Justice*. Oxford University Press.

Table 1: Using different research methods to answer evaluation questions

Research method	Evaluation questions
Analysis of administrative data	• How many clients were referred to the program? • What were the characteristics of clients? • What kind of activities were delivered as part of the program? • Who were the main referring entities?
Data linkage with other organisations	• What are the characteristics of clients referred to the program? • To what extent did the program achieve its intended aims?
Survey	• How satisfied were clients with the program? • To what extent did the program achieve its intended outcomes? • What were the areas for program improvement identified by clients?
Case file analysis	• What were the characteristics of clients referred to the program? • What were the activities delivered as part of the program? • How did client engagement in the program vary over time?
Interviews and focus groups	• Was the program implemented as intended? • What were the main implementation barriers? • What is working well in the program? • How could the program be improved?
Observations	• What were the activities delivered as part of the program? • How was the program delivered on a day-to-day basis? • What was the nature of relationships between clients and service providers?

Administrative data

Administrative data collected by service providers

Administrative data can include information collected by service providers. These data are often by-products of the process of delivering the program and can include:

- referral information provided by referring agencies
- case management information
- registration information and the outcomes of risk assessment processes undertaken by service providers
- attendance records at activities

- information about outputs from activities (e.g., case management plans).

Administrative data collected by service providers can be a good source of information, because they are low cost (i.e., already collected through business as usual), covers the entire client population, and is (generally) systematically collected, stored and used for program operations and management decisions.

How easily useable this information is for evaluation will depend on a number of factors, including:

1. The system used to collect information: The systems used by services to collect this information can range from hard-copy paper files, through to sophisticated case management software programs.
2. The consistency of information being collected (e.g., is everyone using the same definition of 'engagement'?). If service providers are using different rules when they input information into administrative data systems, these datasets will be messy and not easily analysed.

Guidelines 3-5 provide detailed information about how you can improve the useability of your administrative datasets for evaluation and monitoring activities.

Why collect administrative data about what your program is currently doing?

Collecting data about the inputs, activities and outputs associated with your program is important for several reasons. First, it helps you and funding agencies understand the mechanics of your program and what is and is not working well.

Second, it helps you and the funders understand the characteristics of the people who are engaging with your program. This is not only to ensure you are meeting your target client base, but also to identify whether there is unmet need within the community that may require expanding existing programs, funding new programs, or embedding different frameworks and curricula in existing programs.

Third, it helps contextualise and understand the findings from evaluation studies that might be conducted. There is a lot of emphasis on determining whether a program or service 'works', in meeting its intended objectives. However, there is also a need to understand what works for whom, why and under what circumstances.⁹

Administrative data collected by third-party agencies

To assess the longer-term outcomes associated with your program, it may be necessary to link program-level administrative data with datasets maintained by other services and systems that have contact with the same clients. This is because many longer-term outcomes can take years to achieve, long after a client has exited from your program. Table 2 explains how different data sources can help explain the long-term impacts of your program.

However, there are many barriers to data linkage between organisations. Your service is not expected to conduct data linkage. However, you can:

⁹ Burrowes, N & Needs, A (2009). 'Time to Contemplate Change? A Framework for Assessing Readiness to Change with Offenders'. *Aggression and Violent Behaviour*, 14(1): p.39.

- record relevant client information shared through inter-agency collaboration.
- work with partners to explore whether long-term client outcomes can be measured.
- ensure compliance with the *Privacy Act 1988* (Cth), when sharing or using client data for evaluation

By keeping accurate records and engaging in collaborative data-sharing efforts, your service can help build a clearer picture of long-term program outcomes while maintaining ethical and legal standards.

The table below provides examples of the sources of different types of data that can be used to understand a program's long-term impacts.

Table 2: Examples of data sources to understand the long-term impacts of programs	
Use ____ data	to evaluate the impact of programs on:
Births, deaths and marriages	clients' health and wellbeing, including life expectancy
Child protection	the number and nature of child protection notifications and reunification outcomes
Health	clients' health and wellbeing
Housing	clients' housing status, including breaches and issues related to non-payment of rent
Policing	reoffending and desistance
Social security and taxation	clients' health and wellbeing, including access to welfare payments, employment benefits and taxation records

Case file analysis

A case file is a collection of documents relating to a legal, medical or social work case. Although created for administrative purposes (e.g., as evidence of decision-making), case files can be an excellent source of information for evaluation.¹⁰

Case file analysis can produce both quantitative and qualitative data, depending on what evaluation questions are being asked.¹¹ However, because case file data are collected in 'real time', they are particularly useful for documenting clients' trajectories over time in a cost-effective and non-intrusive way. Case files can be analysed thematically and used in a comparative way, to understand the paths of different groups of clients.

Information is usually extracted from case files, using what is known as a coding framework. This is a document that provides key definitions and guidance for extracting data from the case

¹⁰ H Boxall et al (2020). *Responding to Adolescent Family Violence: Findings from an Impact Evaluation*. Australian Institute of Criminology; H Boxall et al (2023). *Sexual Exploitation in Australia: Victim-Survivor Support Needs and Barriers to Support Provision* Australian Institute of Criminology.

¹¹ S Witte (2020). Case File Analyses in Child Protection Research: Review of Methodological Challenges and Development of a Framework. *Children and Youth Services Review*, 108(104551).

files (e.g., how gender and cultural identity are coded), to ensure consistency in how data are extracted and interpreted over time.¹²

Client surveys

Client surveys ask a systematic set of questions, to collect information directly from clients. Surveys can be administered using written or verbal methods.

Like any research method, client surveys have strengths and weaknesses. They allow you to collect both quantitative (closed questions) and qualitative data (open-ended questions). They can be administered to large groups of people quickly and, because they are structured, this makes for easier analysis. Surveys can cover a broad range of topics and outcomes. Depending on the delivery method, they can be efficient and cost-effective.

While surveys are a useful way to collect standardised data from many clients, they also have some important limitations:

- **Limited flexibility** – Most survey questions offer fixed response options, which may not fully capture a client's experience or perspective.
- **Question design matters** – Poorly worded questions can confuse clients or lead to biased results. Each question must be clear, unambiguous, and relevant to the outcome you're trying to measure.
- **Lack of depth** – Surveys are not well suited to exploring complex issues or getting detailed feedback about a client's experience.
- **Barriers for low-literacy groups** – Those with low literacy or limited English (which is very common among criminal justice clients) may struggle to complete written surveys, unless they are supported verbally—something that takes time and resources.
- **Risk of bias when staff administer surveys** – If a familiar staff member conducts the survey, clients may be more likely to give more positive responses to 'please' the staff member. Different staff may also ask or interpret questions in slightly different ways. For this reason, staff who administer surveys should receive training to ensure they deliver questions consistently and record answers accurately.

Surveys in evaluation are typically used to understand:

- the current status or situation of a program
- program outcomes.

Asking the same questions at different time points (e.g., before and after an intervention) can help measure the changes that may be related to a program or service. You can then supplement the survey with other measures, to get a fuller picture of the intervention.

See Guideline 6 for information on how to develop a client outcomes survey, which can be used to measure changes that clients experience because of your program.

Interviews

Interviews are a very well-known method of obtaining detailed information from clients and other stakeholders. **Client** interviews can be a good way of collecting detailed information

¹² Ibid.

about how a program is received by the people it is designed to directly affect (see Guideline 5).

It is also important to gain information from a wide range of other people with relevant information (**stakeholders**), when considering whether a program has been implemented effectively and what kinds of outcomes it has achieved. This can include public servants, politicians, staff of service delivery organisations, families of clients, interest and lobby groups or members of the public.

Some types of interviews are useful for revealing information that the research or evaluation team has not anticipated.

The most common types of interviews are:

- **Unstructured:** an informal, conversational interview that covers broad topics and themes, rather than standard questions. These are good for deriving and exploring initial themes for later research, but a trained evaluator may be required to analyse unstructured content. This will also increase the cost of interviews.
- **Structured:** similar to a survey, these use a standardised set of questions, asked in a fixed order, some of which may have pre-set answer options (e.g., multiple-choice). All questions are asked in each interview. Structured interviews are useful for generating quantitative datasets.
- **Semi-structured:** these have a mixed framework of general themes and pre-established questions, but the questions are adapted to the individual context of the interviews. Interviewers are free to ask follow-up questions or pursue new lines of questioning, as appropriate.

If the participant gives permission for audio-recording, the interviews can later be transcribed and the transcripts analysed (often thematically).

Focus groups

Focus groups are a type of group interview, designed to explore multiple people's views on a topic in depth, in a format in which differences and disagreements can be aired and explored. They tend to have about 6-12 participants and last around 1.5-2 hours. Focus group moderators need skill in facilitating productive, civil conversations and managing group dynamics, while ensuring that all questions and issues are addressed in the allotted time frame.

Observations

There can often be a gap between what people say they do and what they actually do. Therefore, observed behaviour can be a useful complement to self-reported data (such as surveys and interviews) and can allow access to different kinds of evaluation information than other external sources such as administrative data.

Observation is a way of gathering data, by having a trained person systematically watching behaviour, events and phenomena (for example, court hearings or community engagement sessions). Observations can be useful for trying to understand an ongoing process, activity, event or situation; understanding physical settings and how they influence programs and

other interventions; gathering information on behaviours and interactions; or when you can't gather information directly from people, through interviews or surveys.

Observations are usually conducted using structured tools, such as recording sheets, checklists and observation guides. Data collection can be overt, where people know they are being observed, or covert, where they do not. Covert observations raise ethical problems, as they conceal from participants the fact that they are being observed. If using an overt method, there can be different levels of interaction and participation involved in the observation (e.g., unobtrusive, limited interaction, full participation). With appropriate consent, audio and visual recorders or cameras can be used to aid with capturing raw data.

It is important to consider whether being observed might influence people to act differently and, conversely, how observer bias might influence findings (i.e., observers see (or record) only what they want or expect to see). Observation is also a time-consuming and expensive method, compared to other data collection methods, and does not necessarily increase understanding of *why* people behave as they do.

Aboriginal and Torres Strait Islander Indigenous evaluation

So far, the research methods and designs we have described in this guideline have prioritised Western ways of knowing and doing. However, there is increasing recognition of the need to embed the voices and experiences of Indigenous peoples, using methods and processes that facilitate the sharing of these knowledges.

Indigenous worldviews

As discussed above, government understanding of evidence currently prioritises Western worldviews. Aboriginal and Torres Strait Islander worldviews vary across Australia, but often adopt a more collective, rather than individual, approach and involve a deep, spiritual connection to Country. For example, Aboriginal philosopher Mary Graham has described two key concepts: “The Land is the Law” and “You are not alone in the world”.¹³ In the development of the current guidelines, we have consulted (and will continue to consult) with Aboriginal and Torres Strait Islander peoples to ensure that they reflect the Indigenous worldviews.

Centring Indigenous worldviews means also acknowledging that there are likely going to be differences in what Indigenous communities believe constitutes ‘good’ or ‘bad’ evidence. What ‘success’ looks like could differ from the standards applied for non-Indigenous communities or government departments.¹⁴ As noted by Campbell, Kelly and Napaljarri, ‘Because Indigenous and non-Indigenous stakeholders operate from fundamentally different worldviews, they don’t necessarily value the same outcomes... This creates an inherent tension in terms of what constitutes success and which metrics can and should be used to measure it (and whether measurable comparable metrics are helpful)’.¹⁵

¹³ M Graham (1999). ‘Some Thoughts about the Philosophical Underpinnings of Aboriginal Worldviews’ *Worldviews: Global Religions, Culture, and Ecology* 3(2): p.105-106.

¹⁴ M Snijder et al., (2015). ‘A Systematic Review of Studies Evaluating Australian Indigenous Community Development Projects: The Extent of Community Participation, Their Methodological Quality and Their Outcomes’. *BMC Public Health*, 15(1): p. 1154.

¹⁵ Campbell D et al. (2020). ‘Australia’s Indigenous Evaluation Strategy’, ANTAR.

Centring Indigenous worldviews may also lead to using different research methods for collecting information about programs and client outcomes. These research methods could include those that are already being used in Indigenous research projects by Aboriginal and Torres Strait Islander and non-Indigenous researchers, such as yarning circles. Alternatively, it could involve working with Indigenous organisations and researchers to develop new ways of collecting data and undertaking performance monitoring and evaluation activities that are relevant and meaningful for Aboriginal and Torres Strait Islander communities.

Resources and frameworks for Aboriginal and Torres Strait Islander evaluation

In the 2020s, several resources have been developed to support evaluation of Aboriginal and Torres Strait Islander services and programs, and for undertaking Indigenous research more generally. These include:

- the Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS) Code of Ethics;¹⁶
- the Guide to Applying the AIATSIS Code of Ethics;¹⁷
- the Framework for Governance of Indigenous Data;¹⁸ and
- the Indigenous Evaluation Strategy.¹⁹

The AIATSIS Code of Ethics

The AIATSIS Code and accompanying Guide set national standards for the ethical and responsible conduct of all Aboriginal and Torres Strait Islander research, across all disciplines and methodologies. These standards apply to all entities undertaking research, including community organisations.

What is Aboriginal and Torres Strait Islander research?

Under the Code, Aboriginal and Torres Strait Islander research is defined as:

...all research that impacts on or is of particular significance to Aboriginal and Torres Strait Islander peoples, *including the planning, collection, analysis and dissemination of information or knowledge, in any format or medium, which is about, or may affect, Indigenous peoples, either collectively or individually*²⁰ (emphasis added).

Aboriginal and Torres Strait Islander research also includes projects where Aboriginal and Torres Strait Islander individuals or communities are not the explicit target population, but the

¹⁶ Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) (2020). *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. AIATSIS.

¹⁷ AIATSIS (2020)., *A Guide to Applying The AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. AIATSIS.

¹⁸ National Indigenous Australians Authority (2024). *Framework for Governance of Indigenous Data: Practical Guidance for the Australian Public Service*. Australian Government.

¹⁹ Productivity Commission (2020). *Indigenous Evaluation Strategy*. Australian Government.

²⁰ AIATSIS (2020). *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. AIATSIS.

research population includes a significant number of Aboriginal and Torres Strait Islander peoples.²¹

Given the over-representation of Aboriginal and Torres Strait Islander peoples in the ACT criminal justice system,²² it should generally be assumed that data collected about people involved in the criminal justice system *is* Aboriginal and Torres Strait Islander research, unless they are specifically excluded from the program or evaluation process.

AIATSIS Code principles

The AIATSIS Code has a core central value of integrity and is underpinned by four principles: Indigenous self-determination; Indigenous leadership; impact and value; and sustainability and accountability (see Figure 3). Self-determination recognises that Aboriginal and Torres Strait Islander peoples should be provided with the opportunity to engage in and with research that has implications for them and their communities.

²¹ Ibid.

²² Australian Bureau of Statistics (2024). *Prisoners in Australia 2024*.
<https://www.abs.gov.au/statistics/people/crime-and-justice/prisoners-australia/latest-release>

Figure 3: AIATSIS Research Ethics Framework²³



Indigenous data sovereignty

The AIATSIS Code of Ethics core principles reflect the Indigenous Sovereignty Principles and ‘recognises the rights of Indigenous peoples to control the use of their data, wherever it is held (governance of data), and the importance of access to data for Indigenous decision making and self-determination (data for governance)’.²⁴

In practice, this means that Aboriginal and Torres Strait Islander peoples should be involved in key decisions about evaluation and research data at all stages of the research process, including the design stage. This includes determining:

- the questions that are asked as part of Indigenous research projects
- how data are collected

²³ AIATSIS (2020). *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. AIATSIS.

²⁴ Ibid.

- how these data are interpreted; and
- how data are stored and then disseminated.

The Australian Government recently developed a framework for governance of Indigenous data, which sets out four guidelines:

- **Guideline 1: Partner with Aboriginal and Torres Strait Islander people** at all stages of the data lifecycle, to ensure their priorities are reflected in data about their communities.
- **Guideline 2: Build data-related capabilities** – Improve the capabilities of public service staff and Aboriginal and Torres Strait Islander partners relating to Indigenous data, across the data lifecycle.
- **Guideline 3: Provide knowledge of data assets** – Develop straightforward methods for Aboriginal and Torres Strait Islander people to know what data are held relating to their interests, its use, and how it can be accessed.
- **Guideline 4: Build an inclusive data system**, including building towards organisational and cultural change to support the inclusion of Aboriginal and Torres Strait Islander people in data governance.

Best practice

In 2020, the Australian Government released the Indigenous Evaluation Strategy.²⁵

‘Centring’ of Indigenous worldviews is at the centre of the Strategy, underpinning the other principles of credible, ethical, useful and transparent. What centring means in practice includes:

- ‘evaluations are undertaken in the areas, and address the issues, that are most important to Aboriginal and Torres Strait Islander people
- Australian Government agencies routinely consider the impacts of mainstream policies and programs on Aboriginal and Torres Strait Islander people
- Aboriginal and Torres Strait Islander people, organisations and communities have the opportunity to decide how they want to be involved in evaluations
- non-Indigenous evaluators have the necessary knowledge, experience and awareness of their own biases to work in partnership with, and to draw on the knowledges of, Aboriginal and Torres Strait Islander people
- evaluation processes strengthen and support the evaluation capability of Aboriginal and Torres Strait Islander people’.²⁶

The following guidance (at Table 3) from the Cultural and Indigenous Research Centre Australia (2013)²⁷ is useful when evaluating justice programs that have been designed for

²⁵ Productivity Commission (2020). *Indigenous Evaluation Strategy*. Australian Government

²⁶ Ibid.

²⁷ Adapted from Cultural and Indigenous Research Centre Australia (2013) *Evaluation of Indigenous Justice Programs Project B: Offender Support and Reintegration: Final report*. Attorney-General’s Department.

Aboriginal and Torres Strait Islander peoples. Mainstream providers should take account of these themes, including the need for culturally appropriate practices. As Aboriginal and Torres Strait Islander people are consistently overrepresented in the criminal justice system, understanding these themes will assist all providers when evaluating mainstream programs which are not designed purely for Aboriginal and Torres Strait Islander peoples (but are nevertheless likely to include many participants who have Aboriginal and Torres Strait Islander backgrounds).

Table 3: Good Practice Themes for Evaluating Indigenous Justice Programs		
Theme	Evaluation focus	Where is this reflected in the guidelines?
What is a good intervention?		
1: Focusing on crime prevention and aiming to reduce over-representation of Aboriginal and Torres Strait Islander people in the criminal justice system	Does the program provide an evidence-based response to intervention and/or is it based on research about what does/does not work, for whom and under what circumstances?	N/A - This evaluation theme is more likely to be answered by an external evaluator through a review of the literature and through stakeholder consultation processes.
What is a good model?		
2: Meeting needs and addressing a service gap	Does the program fill a service gap and meet needs, which otherwise may be inadequately met or neglected in the service system?	Guideline 3 - Client characteristics Guideline 4 - Activities and outputs
3: Culturally appropriate program design	Is the program culturally appropriate, based and implemented on Aboriginal and Torres Strait Islander empowerment, self-determination and community ownership?	Guideline 3 - Client characteristics Guideline 4 - Activities and outputs Guideline 5 - Client feedback
4: Achieving outcomes in line with program intent	Does the program meet its stated aims and objectives?	Guideline 1 - Program logic models Guideline 6 - Outcomes Guideline 7 - Case studies
5: Promoting inclusive community	Does the program sufficiently engage participation of Aboriginal and Torres Strait Islander people in all stages/aspects and is its model responsive to local needs?	Guideline 3 - Client characteristics Guideline 4 - Activities and outputs Guideline 5 - Client feedback

Table 3: Good Practice Themes for Evaluating Indigenous Justice Programs

Theme	Evaluation focus	Where is this reflected in the guidelines?
6: Effective service coordination and collaboration	Does the program provide an integrated response to the needs of participants?	Guideline 4 - Activities and outputs
7: Advocating for systems reform and improving relationships among key stakeholders	Does the program contribute to advocacy and systems reform and raise the profile of the unique needs of Aboriginal and Torres Strait Islander people within the justice system?	Guidelines 3, 4 and 5 support services to collect information that can then be used for system-wide advocacy efforts
What is a well-managed and -delivered program?		
8: Effective governance and management processes	Does the program have well-defined and effective structures of management and governance, with: • results-based management, linking planning functions with monitoring and evaluation and outcomes-focused • stability and continuity of funding and appropriate resourcing levels • strong leadership and skilled, committed and stable personnel?	N/A - This evaluation theme is more likely to be answered by an external evaluator through stakeholder consultation processes.
9: Clear articulation of program intent	Is the program model clear about the program's aims and objectives, and realistic in scope?	Guideline 1 - Program logic models
10: Sustainability over time	Is there evidence of ongoing support and resourcing for the program?	N/A - This evaluation theme is more likely to be answered by an external evaluator through a review of program documentation and stakeholder consultation processes.