

Evidence in Evaluation

TPG 22-22 Policy and Guidelines: Evaluation – Technical Note

June 2023

Acknowledgement of Country

We acknowledge that Aboriginal and Torres Strait Islander peoples are the First Peoples and Traditional Custodians of Australia, and the oldest continuing culture in human history. We pay respect to Elders past and present and commit to respecting the lands we walk on, and the communities we walk with.

Artwork:
Regeneration by Josie Rose



Contents

Acknowledgement of Country	1
Abstract	3
Context.....	4
Types of Evidence.....	5
Qualitative and quantitative data.....	5
Methods	6
Key data collection methods and sources for evaluation.....	6
Triangulation	12

This publication is protected by copyright. With the exception of (a) any coat of arms, logo, trade-mark or other branding; (b) any third-party intellectual property; and (c) personal information such as photographs of people, this publication is licensed under the Creative Commons Attribution 3.0 Australia Licence.

The licence terms are available at the Creative Commons website at: creativecommons.org/licenses/by/3.0/au/legalcode

NSW Treasury requires that it be attributed as creator of the licensed material in the following manner: © State of New South Wales (NSW Treasury), (2023).

Abstract

Technical Note: Evidence in Evaluation	
Background	Evidence is factual information that can be used to assess a proposition, support a claim or inform policy and decision making. The data sources and data collection tools used should provide evidence on whether the initiative is meeting its delivery and performance objectives.
Scope	<p>This technical note:</p> <ul style="list-style-type: none"> • describes quantitative and qualitative data • identifies data sources and data collection tools and discusses their strengths and limitations • provides information on combining data types and collection methods (triangulation) to strengthen the robustness of evaluation findings.
When to use this technical note?	When collecting evidence for ex-post evaluation, or when undertaking triangulation to build and check the accuracy of evidence collected.
Potential implications	Good evidence is based on data that is appropriate, complete, high quality, transparent and accountable.
Keywords	Evidence, Qualitative, Quantitative, Triangulation, Data
Associated resources	<ul style="list-style-type: none"> • For guidance on how to select a subset of individuals or units (the sample) from within a population to collect data from, see <i>Technical note: Sampling strategy</i>. • For an overview of the different outcome evaluation designs, and the strength of evidence they can provide, see <i>Technical note: Outcome evaluation design</i>.

Context

Evidence is factual information that can be used to assess a proposition, support a claim or inform policy and decision making. Evidence can include quantitative data (data can be counted and expressed in numerical terms) and qualitative data (data that is descriptive, and interpretation-based or conceptual). Good governance of evidence is “the use of rigorous, systematic and technically valid pieces of evidence within decision-making processes that are representative of, and accountable to, populations served” (Parkhurst, 2017, p.5)¹.

There are five principles for the good governance of evidence:

1. **Appropriate** evidence, based on the relevance of the evidence to address the concern.
2. **Complete** evidence, based on or synthesised from a complete set of available information.
3. **High quality** evidence, based on high quality data and the use of appropriately rigorous methodologies.
4. **Transparent** use of evidence and methodologies, which supports public and scientific scrutiny.
5. **Accountable** evidence for reporting that informs policy and decision-making, as well as monitoring and evaluation.

Ex-post evaluation (undertaken after an initiative has been implemented) may use quantitative and qualitative data and a range of data sources and collection tools.

The data sources and data collection tools used should provide evidence on whether the initiative is meeting its delivery and performance objectives. The sources and tools selected should be:

- suitable to answer the evaluation questions
- able to provide the data required, as guided by the evaluation design
- feasible, provide value for money, and within the evaluation budget and timeframes.

Triangulation can be used to build and check the accuracy of evidence collected. This involves combining findings from multiple data sources (e.g., literature review and survey).

¹ Parkhurst, J., 2017. *The politics of evidence: from evidence-based policy to the good governance of evidence* (p. 182). Taylor & Francis. Available at: http://eprints.lse.ac.uk/68604/1/Parkhurst_The%20Politics%20of%20Evidence.pdf

Types of Evidence

Qualitative and quantitative data

In evaluation, the data used may be quantitative or qualitative, or a combination of both (*Table 1: Quantitative and qualitative data*). Some data collection methods (for example, surveys) can generate both qualitative and quantitative data depending on the design.

Table 1: Quantitative and qualitative data

Key Points	Quantitative Data	Qualitative Data
Characteristics	<ul style="list-style-type: none"> Count and express in numerical or diagrammatical terms. 	<ul style="list-style-type: none"> Usually describe using words or images.
Example data sources or collection methods	<ul style="list-style-type: none"> Initiative data (for example performance reports), official statistics, surveys, direct observation. 	<ul style="list-style-type: none"> Initiative data (for example meeting minutes), surveys, interviews, focus groups, direct observation, expert opinion.
Sampling ²	<ul style="list-style-type: none"> Usually probability sampling (for example researcher selects participants at random). 	<ul style="list-style-type: none"> Usually non-probability sampling (for example researcher selects participants using subjective criteria or judgment).
Data analysis ³	<ul style="list-style-type: none"> Ranges from basic descriptive analysis to complex statistical analysis (for example, inferential statistics). 	<ul style="list-style-type: none"> Identifies topics, ideas and patterns that come up repeatedly in responses. Organises information into similar groups or categories (such as clusters).
Example uses in evaluation	<ul style="list-style-type: none"> Identify and illustrate trends. Support the generalisation of results. Provide evidence that changes are attributable to the initiative (outcome evaluation). Test hypotheses or theories generated from qualitative analysis. 	<ul style="list-style-type: none"> Detail context and explain the reasons for quantitative data. Provide a holistic and in-depth perspective of an initiative. Identify important themes to test using quantitative methods. Identify unanticipated impacts.
Limitations	<ul style="list-style-type: none"> Can be undermined by data caveats (for example, sample size or poor response rates). Does not provide details on the context or implementation of the initiative. 	<ul style="list-style-type: none"> May not support generalisation of results. Cannot quantify impacts.

² For more information on types of sampling, sample composition and sample size for quantitative and qualitative analysis, see *Technical note: Sampling strategy*.

³ For more information, see Better Evaluation: https://www.betterevaluation.org/en/rainbow_framework/describe/analyse_data

Methods

Key data collection methods and sources for evaluation

The data used to inform evaluation may come from:

- administrative documents and monitoring data for the initiative or larger program (Table 2)
- other relevant available data sets (internal or external) (Table 3)
- additional data collected specifically for the evaluation (Table 4).

Check what data related to the initiative are already collected⁴, what data are available from other sources that could provide key information on trends or context, and what additional data needs to be collected to address gaps in evidence and understanding.

When deciding on the data sources and the collection methods to use, consider:

- relevance to the evaluation questions
- any restrictions in the use or collection and storage of the data, including ethical concerns (see *Workbook IV. Evaluation plan: Manage the evaluation*)
- strengths and limitations
- the availability of skills, resources and knowledge to collect and analyse the data
- evaluation scope, budget and timeframes.

When selecting data sources and data collection tools, consider what strategies need to be implemented to support cultural appropriateness of collection (for example, pilot test questions) and meet ethical standards that ensure the privacy, dignity and confidentiality of participants (see *Workbook IV. Evaluation plan: Manage the evaluation*).

Assess the quality and usefulness of the data, considering:

- Validity – is the data accurately measuring the identified concern?
- Reliability – can the data be replicated consistently?
- Completeness – is there any missing data?
- Precision – do the data sets have sufficient detail?
- Integrity – is the data protected against any bias or manipulation?
- Availability – is the data sets accessible?
- Timeliness – is the data current and will they be available when needed?

⁴ Reference the initiative's monitoring and evaluation framework.

Table 2: Administrative documents and monitoring data for the initiative

Data source	Description	Strengths	Limitations
Initiative data	<p>Available information and routine data, for example:</p> <ul style="list-style-type: none"> • administrative data (for example, delivery and client data) • monitoring data <ul style="list-style-type: none"> — monitoring (as planned under a Monitoring and Evaluation Framework) — performance reporting — benefits realisation management framework • records, minutes, correspondence, memos and reports • initiative documentation (for example, business case and project plan) 	<ul style="list-style-type: none"> • can provide detail about the initiative's implementation • can provide regularly collected quantitative or qualitative data • is a cost-effective way of obtaining data • may point to topics to include in interviews or surveys 	<ul style="list-style-type: none"> • some of the data may only reflect the views of the individuals involved (for example, meeting minutes) • dominance of output information (e.g. number of services provided) and lack of outcome data (e.g. increase in user satisfaction) • some of the data may be incomplete
Outcome budgeting data; State Outcome Indicators and Program Performance Measures	<p>An Outcome Indicator is a measure of effectiveness that can reasonably demonstrate to the public the performance of the New South Wales Government in achieving a State Outcome.</p> <p>A Program Performance Measure (PPM) is a quantitative or qualitative measure of <i>program</i> performance that is used to demonstrate change.</p> <p>The impact of an initiative on State Outcomes and PPMs varies according to the size and purpose of the initiative. A large initiative may have a direct impact on a State Outcome. Smaller initiatives may contribute towards or combine with other initiatives to influence State Outcomes.</p>	<ul style="list-style-type: none"> • draws upon established indicators • supports alignment of evaluation with state outcomes objectives <p><i>Note: performance indicators used within an initiative evaluation will likely be more comprehensive and focused on the detail of activities and their outcomes but should be aligned to the relevant State Outcome Indicators and PPMs.</i></p>	<ul style="list-style-type: none"> • more detail may be required to support higher level indicators • relationship between state outcomes and initiative will need to be investigated

Table 3: Other relevant available data sets

Data source	Description	Strengths	Limitations
Agency/ cluster data	Internally available or published agency or cluster data. Published examples include: <ul style="list-style-type: none"> • NSW Department of Primary Industries Performance data and Insights • Transport for NSW: Transport Performance and Analytics 	<ul style="list-style-type: none"> • utilises existing data (minimises additional activities and costs) • may provide baseline data • may provide information on relevant patterns and trends • may suggest associations or provide evidence of changes that merit further investigation 	<ul style="list-style-type: none"> • measurable changes cannot be attributed to the initiative without investigation • data may not be fit for purpose (for example, may not address the relevant time-periods or cohorts) • data may not identify initiative's participants (even if the cohort of interest is captured) • validity and reliability of data may need to be verified in some cases • access to certain datasets may require authorisation, which may be time consuming
Official statistics	Examples include: <ul style="list-style-type: none"> • Australian Bureau of Statistics • Australian Bureau of Agricultural and Resource Economics and Sciences • Australian Institute of Health and Welfare • Australian Data Archive • National Centre for Vocational Education Research 		
Online data sets	Examples include: <ul style="list-style-type: none"> • Household Income and Labour Dynamics in Australia survey, Melbourne Institute • Illion data registries • NAB Monthly Business Survey • Westpac MI Consumer Confidence • Regional Wellbeing Survey, University of Canberra 		
Other sources	Other sources may include: <ul style="list-style-type: none"> • publications (for example, books, or journals) • online resources (for example, websites and forums) • web data (for example, data generated by users assessing websites) 		

Table 4: Data collection methods for evaluation

Collection method	Description	Strengths	Limitations
Surveys (including customer satisfaction surveys)	<ul style="list-style-type: none"> • Consistent set of questions asked to a large group of people, about their experiences, level of satisfaction, opinions, attitudes and motivations. • Can be designed to provide both quantitative and qualitative data. • Delivered by mail, internet, telephone, mobile or in person. • Initial pilot survey, using a smaller group of participants, can be used to ensure that questions are being interpreted as intended and produce the expected information. 	<ul style="list-style-type: none"> • can be designed to collect data from a large number of participants in a short period of time • can be relatively low cost • can be anonymous • can be easy to analyse responses, since they are standardised 	<ul style="list-style-type: none"> • risk of low response rate, particularly when collecting data from hard-to-reach populations • risk of biased responses (for example, social desirability bias or if participants are those benefiting from the initiative) • requires expertise to design questions, to ensure that questions are framed appropriately in line with the key questions in order to optimise the collection of rich information without discouraging respondents from completing the survey • sample size and composition need to be considered when generalising results
Expert opinion	Opinions sought from leading professionals in relevant fields.	<ul style="list-style-type: none"> • targets key expertise 	<ul style="list-style-type: none"> • depends on credibility and expertise of participants • risk of biased responses (for example, subject to the experts' own biases and opinions)

Collection method	Description	Strengths	Limitations
Interviews	<p>Questions asked about participants' experiences, opinions, attitudes and motivations, using:</p> <ul style="list-style-type: none"> • Structured interviews – use the same set of questions for all interviewees (useful when there are several people conducting interviews) • Semi-structured interviews – use an interview guide or checklist, rather than fixed questions (ensures comparable data are collected, while giving flexibility to collect extra information) • Unstructured interviews – use questions tailored to each participant 	<ul style="list-style-type: none"> • can produce rich data and descriptions • can identify trends, themes or issues that may merit further examination (for example, with quantitative data) • provides opportunity for in-depth questioning in response to areas of interest that emerge 	<ul style="list-style-type: none"> • risk of biased responses and data collection (for example, social desirability bias, or interviewer confirmation bias⁵) • findings are not generalisable beyond the interviewed population • requires skilled and unbiased interviewer • can be time consuming to conduct interviews and analyse responses • sample size and composition need to be considered when generalising results
Focus groups	<ul style="list-style-type: none"> • Data collected from several participants during shared sessions. • Be aware of the characteristics of participants (for example, gender, age, organisational position etc,) when selecting the group. • Semi-structured approach is recommended. • Involve expertise in workshop design and facilitation. 	<ul style="list-style-type: none"> • can produce rich data and descriptions • can identify trends, themes or issues that may merit further examination (for example with quantitative data) • provides an opportunity for exploration and development of ideas through group discussion 	<ul style="list-style-type: none"> • group settings may inhibit or influence opinions (for example groupthink bias) • some individuals may impact the group dynamics • not suitable for collecting sensitive information • requires skilled facilitator • findings are not generalisable beyond the interviewed population • can be time consuming to conduct focus groups and analyse responses

⁵ These can happen when participant responses are shaped by what they think is correct or acceptable or when interviewers only include data they consider relevant.

Collection method	Description	Strengths	Limitations
Direct observation	<ul style="list-style-type: none"> • Field notes that can provide quantitative and qualitative data. • Observers unobtrusively and systematically record encounters within an initiative setting (includes case studies). 	<ul style="list-style-type: none"> • provides rich, detailed descriptions of observations and their context • can identify issues not reported in interviews 	<ul style="list-style-type: none"> • requires expertise, to ensure that data are collected and recorded uniformly • participants may change their behaviour if they know they are being observed (Hawthorne effect) • can be resource intensive • requires opportunity for unobtrusive observation • potential observer biases • findings are not generalisable beyond the observed population

Triangulation

Triangulation is a way of combining data or findings from multiple sources, that investigate the same subject, to build and check the accuracy of evidence. It is good practice in evaluation to triangulate different methods or data sources (see *Table 5*)⁶⁷. Triangulation can be used at different levels in a methodology: methods, measures, and data triangulation, as well as investigator and theory triangulation.

Table 5: Types of triangulations

Types of triangulations	Description
Data triangulation	Use different data sources (for example, different locations or participants).
Methods triangulation	Use multiple methods to investigate the evaluation questions, or use one method to enhance or clarify the results of another (for example, surveys and interviews).
Investigator triangulation	Use more than one investigator, to decrease bias or confirm findings.
Theory triangulation	Use multiple theories or hypotheses, to provide different perspectives from which to investigate a question.

For example, using methods triangulation in a process evaluation may involve using results from interviews or focus groups to confirm the information collected from administrative documents. Using methods and theory triangulation in an outcome evaluation may involve using qualitative analysis to confirm the causal attribution identified through experimental or quasi-experimental designs.

Triangulation can be used to:

- offset the limitations and biases that can arise from using a single design, method or source
- provide a more comprehensive overview and understanding of the initiative
- address complex questions
- improve the validity and credibility of evaluation findings
- understand unexpected findings
- assist in interpreting the evaluation results
- balance different perspectives and generate new hypothesis, where data produces different or contradictory findings
- build data collection quantity or quality.

⁶ When combining information from different administrative or survey sources to provide new datasets for research purposes, that involve Commonwealth data and are considered 'high risk', an accredited Integrating Authority must be used. For more information see: <https://toolkit.data.gov.au/data-integration/roles-and-responsibilities/integrating-authorities.html> (accessed 6 June 2023).

⁷ https://www.unaids.org/sites/default/files/sub_landing/files/10_4-Intro-to-triangulation-MEF.pdf