NSW Health

www.health.nsw.gov.au

POPULATION HEALTH GUIDANCE SERIES

Qualitative methods for health policy and program evaluation

A guide

NSW GOVERNMENT

NSW Ministry of Health
1 Reserve Road
ST LEONARDS NSW 2065
Tel. (02) 9391 9000
Fax. (02) 9391 9101
TTY. (02) 9391 9900
www.health.nsw.gov.au

Produced by:
Centre for Epidemiology and Evidence

The NSW Ministry for Health acknowledges the traditional custodians of the lands across NSW. We acknowledge that we live and work on Aboriginal lands. We pay our respects to Elders past and present and to all Aboriginal people.

Further copies of this document can be downloaded from the NSW Health website www.health.nsw.gov.au

SHPN (CEE) 250943
ISBN 978-1-74231-285-9

January 2026

Contributors to the development of the guide:
Anne Grunseit in collaboration with Melanie Andersen, Robyn Newson and Ludmila Ovchinikova from the NSW Ministry of Health.

Acknowledgements
This guide was commissioned by the Centre for Epidemiology and Evidence, NSW Ministry of Health and was prepared by Anne Grunseit. The authors would like to thank Michelle Dickson, Darkinjung/Ngarigo, and Megan Williams, Wiradjuri, for their review of and input into the guide.

Suggested citation:
Centre for Epidemiology and Evidence. Qualitative methods for health policy and program evaluation: A guide. Evidence and Evaluation Guidance Series, Population and Public Health Division. Sydney: NSW Ministry of Health, 2026.

Contents

1. Executive summary

This guide has been designed to give users, who may be unfamiliar with qualitative research approaches, an overview of the most common qualitative methods used in the evaluation of health initiatives,* along with good practice principles and examples. It also includes a checklist to assist health policy and program staff who are conducting evaluations to assess the completeness and appropriateness of evaluation proposals and reports which use qualitative methods.

Using qualitative methods in evaluation

Qualitative approaches in evaluation can include methods such as focus groups, in-depth interviews, observation, or analysis of existing documents or media to explore complexity, meaning, relationships and patterns. Words, and sometimes images, are the unit of analysis. Qualitative methods are well-suited for examining topics such as the attitudes, behaviours and lived experiences of people or groups of people, and are useful when it is important to understand the context within which something occurs. They can be informative in evaluation to answer 'what', 'why' and 'how' questions. In contrast, 'how many' or 'how much' questions are best answered using quantitative techniques. Notably, findings from qualitative evaluation studies are not intended to be generalised beyond the context in which the study was carried out; the purpose of qualitative research and evaluation is to provide in-depth explanations and meanings rather than generalisable findings.

Qualitative approaches in evaluation are often employed together with quantitative approaches (mixed methods) so findings can be triangulated (combined and compared) to provide more comprehensive answers to evaluation questions. Qualitative approaches may also be used to address separate evaluation questions to those addressed through quantitative approaches or as a stand-alone approach.

Rigour in qualitative evaluation

High quality qualitative evaluation studies address a clear evaluation question(s) using rigorous methods, including appropriate sampling and data collection, and systematic data analysis and interpretation. Sufficient description of the methods employed should be provided in proposals and reports so their relative strength and usefulness can be assessed. Data collection decisions (e.g. choices between interviews, focus groups, open-ended survey questions) should consider the richness and volume of data appropriate for the analytic method being used (e.g. thematic analysis, content analysis). Sampling should consider the population group and approach best able to answer the evaluation questions; the main aim being to sample data sources with sufficient range and depth to adequately understand, explain or describe the topic of interest (i.e. achieve data adequacy). Analytic methods may provide descriptive categorisations of data, or more interpretive insights (e.g. thematic analysis), however, evaluators should demonstrate an understanding of the difference and accurately describe the analytic approach adopted. Interpretation should involve the appropriate integration of quantitative and qualitative findings (if relevant) and findings presented should be credible and compelling, with clear relationships evident between the data and the conclusions drawn.

Ethical and cultural considerations

As with any form of evaluation, the ethical, personal, social, and cultural factors which may impact participants and stakeholders should be considered when using qualitative methods. Evaluations involving Aboriginal people(s)** should be conducted in accordance with relevant guidelines, including the Aboriginal Health and Medical Research Council (AH&MRC) of NSW Ethics Guidelines. These include requirements about Aboriginal governance,

* In this guide the word 'program' is used interchangeably with 'initiative'. The *NSW Treasury Policy and Guidelines: Evaluation (TPG22-22)* define an initiative as a program, policy, strategy, service, project, or any series of related events. Initiatives can vary in size and structure; from a small initiative at a single location, a series of related events delivered over a period, or whole-of-government reforms with many components delivered by different agencies or governments.

** In this Guide, Aboriginal and Torres Strait Islander people are referred to as Aboriginal people in recognition that Aboriginal people are the original inhabitants of NSW.

involvement of Aboriginal people throughout the evaluation process, Aboriginal community control, and cultural sensitivity. The appropriate use of Indigenist research methodologies and Indigenist qualitative data collection methods can improve the experiences of participants, improve the quality of data collected, and improve the relevance and utility of evaluation findings. Indigenist research approaches adopt a strengths-based stance and centre Aboriginal ways of knowing, being, and doing.

Well-conducted qualitative evaluation studies can provide unique and valuable insights to inform the development and implementation of health initiatives, and to explore initiative impacts. Ensuring a good understanding of what qualitative methods involve and where they fit in the evaluation toolkit is fundamental to their appropriate use.

2. About this guide

NSW Health is committed to ensuring that evaluation conducted within the NSW Health system is rigorous and enhances the evidence base for public health decision-making. This brief guide is designed to support health policy and program staff to engage an independent evaluator[†] and/or appraise evaluations that use qualitative methods.

2.1 Scope and purpose

This guide has been designed to give users, who may be unfamiliar with qualitative research approaches, an overview of the most common qualitative methods used in evaluation of health initiatives, along with good practice principles and examples. Policy and program staff bring a wealth of applied expertise to the process of evaluation and their involvement is key to ensuring good quality and useful evaluations. The information in this guide aims to build users' confidence in knowing 'what to look for' when managing evaluations involving qualitative methods. It is not intended to provide a 'how to' or comprehensive set of technical instructions for conducting qualitative research-based evaluations, nor is it exhaustive. Readers interested in more technical information can consult texts such as Liamputtong (2019)[1] or Leavy (2014),[2] or other references cited in this guide.

2.2 Structure

Section 3 provides an overview of qualitative methods and the types of questions that are suitable for this set of techniques. Sections 4–7 cover qualitative data collection, sampling, analysis, and reporting. At the conclusion of each section the key implications are outlined for engaging an independent evaluator, assessing proposals,[‡] and assessing reports. Examples are used to illustrate the application of concepts in practice, including many Australian studies. Definitions for technical terms are provided throughout as text links with a hover-over function and at the end of the guide (Section 10). A checklist to support the appraisal of evaluation proposals and reports with regard to qualitative methods can be found in Section 8. This checklist is designed for use by health policy and program staff and is adapted from qualitative research checklists developed for academic users.

2.3 Other evaluation guidance

This guide is intended for use in conjunction with other publications that provide broader guidance about the planning, conduct, reporting, and use of evaluations. Key documents from the Population Health Guidance Series include:

- *Planning and Managing Program Evaluations: A Guide*[3]
- *Preparing and Appraising Evaluation Reports: A Checklist*[4]
- *Study Design for Evaluating Population Health and Health Service Interventions: A Guide*, which includes information about quantitative evaluation approaches.[5]

In addition, the *NSW Treasury Policy and Guidelines: Evaluation (TPG22-22)*[6] set out mandatory requirements, recommendations and guidance for NSW General Government Sector agencies when planning and conducting evaluations. The accompanying evaluation workbooks and resources contain information to support evaluation planning and conduct.[7]

A resource list including resources related to ethical review and conducting evaluations with Aboriginal people(s) can be found in Section 9.

[†]An independent evaluator may be an individual or group external to the policy team managing the program whether internal (e.g. evaluation team within the Ministry of Health) or external (e.g. consultant or academic group) to the program delivery agency. Engaging an independent evaluator is important where there is a need for special evaluation expertise and/or where independence needs to be demonstrated.

[‡] Noting that the items for proposals may also serve to assess evaluation plans.

3. Introduction to qualitative methods in evaluation

Well-conducted evaluation studies using qualitative methods can provide unique and valuable insights to inform the development and implementation of health initiatives and explore initiative impacts. For example, qualitative methods can help us to better understand patients' experiences, interprofessional dynamics and the factors affecting the success of health service delivery. Qualitative methods can be an informative companion to quantitative methods in evaluation but also have merit as a stand-alone approach. They make important contributions to health service evaluations that cannot be captured as well through other techniques. Ensuring a good understanding of what qualitative methods comprise and where they fit in the evaluation toolkit is fundamental to their appropriate use.

3.1 What are qualitative methods?

Qualitative methods encompass a large variety of data collection and analysis techniques.[8] What they have in common is that where quantitative methods use numbers, qualitative methods most commonly use words, and sometimes images. Qualitative data are most often drawn from interview and focus group transcripts, policy documents, open-ended survey questions, observational notes, or media articles and advertising images. In analysis, qualitative methods involve organising the data and systematically looking for patterns and principles to interpret and explain social phenomena.

Box 1: Trustworthiness and rigour in evaluations using qualitative research

Bias in research arises from deviations from the truth during any part of the research process (e.g. data collection, data analysis, interpretation or publication), causing distorted results and potentially false conclusions to be drawn.[9] Both quantitative and qualitative research methodologies can be prone to bias and employ different strategies to mitigate bias and ensure trustworthiness and rigour.

Quantitative research is based in the notion of scientific objectivity. From this lens, a study should not be biased by any 'subjective' factors such as an evaluator's personal interpretations. It is considered important for the evaluator to remain detached from the research process; for subjectivity in measurement, analysis and reporting to be limited; and for the influence of contextual factors to be minimised. Steps are taken in high quality quantitative research to address various forms of bias (e.g. selection bias, measurement bias) in order to maintain objectivity and to achieve valid and reliable results that can be replicated and are generalisable.[10] Even so, quantitative evaluators' views and the social and political climate in which research is conducted can also influence several aspects of quantitative research, including the topic studied, the questions asked, the analytic approach, and how findings are interpreted and presented.[11-13]

Qualitative research, in contrast, aims to provide in-depth explanations and meanings that necessarily rely on exploring 'subjective' materials such as the opinions and perceptions of study participants. The context in which the study occurs is considered an integral part of the analysis and, rather than being detached, the evaluator must play a direct and conscious role in choosing questions, collecting data, and interpreting findings. In addition, the findings of qualitative research are not intended to be generalised beyond the context in which the study was carried out.

Trustworthiness and rigour in qualitative research are conceptualised as credibility, dependability and confirmability of findings, achieved through systematic methodology and transparent reporting.[14] Evaluator(s) should be transparent and reflexive about their own preconceptions, relationship dynamics and analytic focus, and acknowledge the role they play during the research process (i.e. data collection, analysis, reporting etc.).[15] Procedures such as declaring evaluators' disciplinary backgrounds, recording and transcribing data, providing detailed descriptions of the process of analysis, and checking interpretation of findings with study participants (member checking) are also used to enhance rigour.[16] Analytical methods such as double coding of transcripts by two or more evaluators, and assessing inter-rater reliability, can also augment rigour and trustworthiness.

The checklist in Section 8 includes items to assess rigour in proposals and reports using qualitative methods.

3.2 What evaluation questions are suitable for qualitative approaches?

Qualitative approaches are well-suited for examining topics such as the attitudes, beliefs, behaviours and lived experiences of people or groups of people, and when it is important to understand the context within which something occurs. They can be informative in evaluation to answer 'what', 'why' and 'how' questions. In contrast, 'how many' or 'how much' questions are best answered using quantitative techniques. Tables 1 and 2 include examples of evaluation questions suited to qualitative or quantitative approaches, respectively.

Box 2: Selecting appropriate evaluation questions

Evaluations are only as good as the questions that they ask. Deciding on the right questions is not always easy or straightforward. Those managing evaluations need to carefully consider the purpose of the evaluation and which questions will best serve that purpose. It is the questions which should dictate the methods used rather than the reverse. Setting out clear and evaluable questions enables appropriate methods to be chosen to ensure an evaluation provides useable findings. For further information on generating evaluation questions, refer to *Planning and Managing Program Evaluations: A Guide*[3] and the NSW Treasury's *Evaluation Workbook III. Evaluation plan: Design the Evaluation*.[17]

Table 1. Example questions for qualitative approaches

Qualitative question examples
Purpose: Hypothesis generating, exploratory, reveals mechanisms of process and impact and meaning
How was component X of a health initiative implemented?
What did this aspect of the initiative mean to participants?
What contextual factors help to explain people's experience of the initiative?
Do groups X and Y have different understandings of an initiative?
How do understandings of an initiative change over time?
To what extent did participants engage with the initiative?
How well did the initiative meet participants' needs?
Did participants perceive the initiative to be useful and/or impactful?

Table 2. Example questions for quantitative approaches

Quantitative question examples
Purpose: Hypothesis testing, sample to population conclusions, causal inference and enumerating
How many people participated in component X of the initiative?
What is the average and range of the outcome in population X?
Did the initiative change outcome X?
What factors moderated/mediated the effect of the initiative?

3.3 Qualitative methods for different types of evaluation

Several different types of evaluation for health initiatives may be appropriate depending on where the initiative is in its development cycle.[3,6] Formative evaluation may be undertaken in the early stages of program design and implementation to inform decisions about initiative improvement. Process evaluation gauges how well an initiative has been implemented, and explores factors such as initiative reach, acceptability and appropriateness. Outcome evaluation (sometimes also called impact evaluation) examines the short, medium, and longer-term effects of an initiative.

Qualitative methods are appropriate for all these evaluation types. They are useful during initiative planning as well as for assessing its implementation and understanding its effectiveness. Table 3 gives examples of real-world studies using qualitative methods for different evaluation types.

Table 3. Example qualitative methods for different evaluation types

Evaluation type	Evaluation objective	Data collection
Formative Cranney et al, 2018[18]	To identify engagement strategies to increase the Chinese community's NSW Get Healthy Service participation.	Semi-structured interviews with community stakeholders (n=16).
Process Manby et al, 2022[19]	To explore healthcare workers' perceptions and attitudes towards vaccines and the COVID-19 vaccination program in the UK.	In-depth, semi-structured telephone interviews with frontline healthcare staff (n=24); analysis of COVID-19 vaccination policies and guidance documents.
Outcome McGill et al, 2020[20]	To determine how participants in the Healthy Weight for Life Long Term Maintenance Program account for their engagement with a weight loss maintenance program and the role of the program in their weight management.	Interviews with 32 participants in a private health insurance run weight loss maintenance program.

3.4 Using qualitative methods in mixed methods evaluation

Where qualitative methods are used in combination with quantitative methods, this is called 'mixed methods'. There are many ways these two approaches can be used together in terms of:

- how they are sequenced (what comes first and how each might inform the execution of the other)
- the relative weighting given to each method (whether one method is more dominant or subordinate in the evaluation than the other)
- whether qualitative and quantitative findings are integrated, and if so, at what stage of the analysis or interpretation.

It is considered good practice in evaluation to combine or 'triangulate' data from different sources (e.g. qualitative and quantitative methods) to see whether the findings from one data source are corroborated by the other. Consistency across data sources gives evaluators and other stakeholders confidence that the conclusions being drawn are valid, noting that inconsistencies are not necessarily problematic and can in some cases help to clarify evaluation results. Therefore, triangulation is often cited as a reason for using a mixed methods approach. However, there are many other reasons why an evaluator might deploy both qualitative and quantitative methods within the one evaluation. For example, using qualitative methods to explain patterns observed in a quantitative analysis; using qualitative methods to develop hypotheses which can then be tested quantitatively in a population representative sample; or simply answering different evaluation questions in a complex evaluation.[21] Importantly, qualitative methods are not merely supplementary to quantitative in mixed methods evaluation; the relative role and importance of each depends upon the overarching evaluation questions and purpose.

Figure 1. Implications for the use of qualitative methods in evaluation

Implications for engaging an independent evaluator

- Provide clear and executable evaluation questions, to which methods can be matched
- Consider whether to specify the use of qualitative or mixed methods
- Consider which evaluators have the necessary skills and experience in qualitative methods

Implications for assessing proposals

- Qualitative methods are well-matched to relevant evaluation questions, the type of evaluation, target population and current knowledge gap
- Data types proposed are appropriate to target population and qualitative methods
- If mixed methods are suggested, the purpose for their use is clearly articulated and compelling

Implications for assessing reports

- All aspects of qualitative methods are clearly reported
- Methods used are suitable for answering qualitative evaluation questions
- Includes clear explanation of the integration of qualitative and quantitative data in mixed methods evaluation, which is also reflected in reporting

4. Qualitative data collection

The data collection process in qualitative evaluation critically influences its quality and utility. The most important aspect of selecting a data collection method is whether that approach will yield data appropriate to answering an evaluation question. Although data collected should be the best it can be within available resources (time, participants, budget, capacity), there are often practical constraints to 'ideal' data collection.

Data collection decisions should consider the richness and volume of data appropriate for the analytic method being used (see Section 5). For example, where data are collected as open-ended questions in a survey, the range of suitable analytic methods is narrower because the richness and volume of such data are often limited compared with longer form techniques such as focus groups, interviews and yarning described below. Data collection decisions should also consider any ethical, personal and cultural factors which may impact the people from whom data is being collected (see Box 3). All data collection should be conducted in accordance with relevant Human Research Ethics Committee requirements.

Box 3: Ethical practice when using qualitative methods in evaluation

Although ethical practice is important for all types of evaluation, there are some considerations that are particularly relevant to qualitative data collection. For example, qualitative data collection can be time consuming for participants and therefore evaluators should reduce potential burden (a type of harm) by only collecting data that is relevant to the evaluation questions, not merely because it is interesting. It may also be appropriate to compensate participants for the time they spend contributing to the evaluation.

In addition, qualitative data collection involves the use of open-ended and dynamic questioning so it is not always possible to know in advance what a participant may reveal and what implications might arise as a result. Therefore, there may be more risk of psychological harm or breach of privacy for participants compared to the more circumscribed process of a survey. Prior consideration should be given to managing any psychological distress and potential privacy concerns during qualitative data collection. Particular care should be taken where the topic being discussed is sensitive, as is the case with many health-related investigations.

Further, the needs of the specific population/cohort participating in the evaluation should be considered. Data collection undertaken with populations who may experience health or social disadvantage (e.g. Aboriginal people(s), LGBTIQA+ people) requires management of sensitivities and should not compound disadvantage. Insensitivity to those who are being asked to share their insights and knowledge is not only contrary to ethical practice[22] but may actively cause harm to participants. Aside from this, inappropriate techniques will yield lower quality data as participants withhold information. Special consideration should also be applied where there may be issues regarding consent or understanding (e.g. children, people from culturally or linguistically diverse backgrounds, or people with intellectual disabilities).

Further resources to decide if review by an ethics committee is needed can be found in Section 9, noting that evaluations involving or reporting findings about Aboriginal people as a group need to follow Aboriginal Health and Medical Research Council (AH&MRC) guidelines[23] and the NSW Health *Quick Guide on Undertaking Appropriate Aboriginal Health Research*.[24]

4.1 Data sources

As noted in the previous section, words are the primary type of data underpinning qualitative approaches. Sources of those words can be existing documentation – essentially any text-based document such as policies, media articles, social media posts, webpages, and clinical case notes that already exist. Other sources may be generated through the evaluation process, such as transcripts from interviews and focus groups, open-ended survey questions, observational and field notes, and journals. An evaluation may use one or more of these data sources. For example, an evaluation of the Get Healthy at Work initiative undertook focus groups with employees, and interviews with business key contacts and service providers.[25]

As with all data collection, using existing data often has the advantage of requiring fewer resources and reducing participant burden. Considering what data sources are already available is thus often a useful first step. Qualitative data collection can be as straightforward as gathering an appropriate sample of text documents or images to underpin an analysis (see Section 5). In cases where new data need to be generated during an evaluation, there are a number of commonly used approaches.

4.2 Common qualitative data collection methods

There are many data collection methods used in qualitative approaches. This guide will briefly describe the purpose, advantages and disadvantages of the most common methods used in the evaluation of health initiatives. As for all forms of data collection, data collection tools (e.g. interview guides or questions, content analysis data extraction tools, data collection protocols) should be reviewed by those managing the evaluation prior to data collection commencing to ensure they are fit for purpose.

4.2.1 Focus groups

Focus groups feature frequently in evaluations of health initiatives. In general, between six and eight people are brought together in a room (or a videocall) and interviewed simultaneously on a specific topic. Having fewer than six people can limit the advantage of interchange between participants; having more than eight can limit the willingness or ability of participants to offer their opinion or lead to side conversations between subgroups of participants. Data collection is usually guided by a combination of a pre-prepared discussion topic guide, an independent facilitator or evaluator, and the natural conversational pathway the group takes. Group composition should avoid potential power inequalities (e.g. managers and junior staff) or strongly opposing ideological positions (e.g. drug users and anti-drug campaigners) between group members as this can stifle disclosure.

Importantly, focus groups are not about efficiency where many opinions are gathered at the same time. Their value over one-on-one interviews is that participants prompt each other and more can be revealed when people propose, defend, and even change, their position on a certain topic. For example, a participant may take the 'position' that people who develop chronic disease should not receive public health assistance because they are responsible for their condition through 'choosing' an unhealthy lifestyle. Other participants may contest this logic, perhaps arguing that the food marketing and purchasing environment influence that person's choices. The first participant may then respond to those arguments by interpreting them in a way which either confirms their position or concedes their merit through reformulation of their position. The rationales put forward and how participants interpret them can provide additional insights into a topic.

Table 4. Advantages and disadvantages of focus groups

Advantages	Disadvantages
Unique insights gained through the interactions of group members and immediate feedback on people's standpoints	The practicality of organising participants to meet at the same time
May encourage disclosure on stigmatised issues where participants share the same situation[14]	Limitations in the range of topics which can be explored due to privacy and/or sensitivity and risks associated with information disclosure to other participants in the group
Opportunity to test new concepts (e.g. messaging for a health campaign) among a diverse group	Group dynamics are hard to predict and may work against yielding useable data (e.g. when one person dominates a discussion, unanticipated power imbalances etc.)

4.2.2 Interviews

Interviews are another common qualitative technique used in evaluation. They can be conducted one-on-one, or occasionally as 'group' interviews of two or three respondents. The latter is distinct from focus groups in that a pair (or less commonly, a trio) may be interviewed simultaneously because they co-share, for example, a position central to the interview topic (such as co-managers of a health initiative). Qualitative interviews are also distinct from interviewer-administrated surveys where a questionnaire of close-ended questions is administered by an evaluator (as opposed to respondent self-completion).

An interview schedule may vary in terms of the specificity of topics (broad or narrow), the order of topics (fixed or flexible), and the license to depart from those topics (e.g. greater latitude may be given if the evaluation question is more exploratory). Interview schedules may therefore be described as structured, semi-structured, or unstructured. Irrespective, an interview should strike a balance between allowing the interviewee(s) to reveal information not anticipated by the evaluator, obtaining information pertinent to the evaluation question and respecting participant burden. Typical interview length varies between 30 to 90 minutes and can take place over several sessions, although practical considerations usually limit interviews to one session.

Table 5. Advantages and disadvantages of interviews

Advantages	Disadvantages
Can explore a topic in depth	Can be resource intensive (time and financial)
Can explore sensitive topics	May cover only a limited range of experiences
Conducive to rapport building between interviewer and interviewee, which can facilitate disclosure	
May obtain full and contextualised narrative of a health topic (e.g. comprehensive description of an encounter with a health service)	

Box 4: Recording and transcription versus note taking

When planning data collection through methods like interviews, focus groups, or yarning, evaluators must decide whether they will, with the permission of interviewees, audio/video record and subsequently transcribe the conversations or take notes. This is not an insignificant decision. The two differ greatly in terms of the detail which can be accurately captured and what data will be available for analysis.

It is preferable to record and transcribe data collected from participants with their permission rather than take notes for several reasons. Audio/video recordings can be transcribed verbatim, while notes will only provide an overview of the discussion. Notes are also more susceptible to bias as the notetaker chooses what is recorded and how to summarise it. Having a verbatim record of participant responses allows the evaluator to draw on the richness and nuance of people's own words in a way that notes cannot. Often the expressions and language chosen by participants is central to understanding people's experiences and explanations of their thoughts and behaviour. Other aspects such as hesitations or laughter can be easily linked to speech to give further insights into the data. From a practical perspective, if the resources or circumstances of data collection do not extend to a notetaker as well as an interviewer, the interviewer may miss parts of the interview when trying to capture people's words. Transcripts also provide a complete record to share with participants for member checking purposes.

4.2.3 Observational and field notes

Observational notes are taken at the time of observing practices or behaviours of interest in their natural context; they can be structured or unstructured. They are often used in evaluation studies where an evaluator may be embedded in a social or cultural group either as participant (e.g. a nurse working in an intensive care unit) or non-participant. Content can be as diverse as describing hygiene behaviours of clinicians in a hospital ward, negotiation between users of outdoor gym equipment in a public park, and cultural practices within a community. In contrast to full transcripts of formal interviews, observational notes are mostly the words of the observer-evaluator, although they may contain fragments of quotes or interactions with the observed group, especially to document significant phrases or terminology.

Field notes are similar to observational notes in that they record the evaluator's thoughts rather than the words of an interviewee. In contrast to observational notes, field notes are primarily used to support the evaluator's recall but may also be used as data in an analysis. They are often taken as an adjunct to interviews to document contextual aspects such as the evaluator's impressions of the interviewee, the setting and perhaps commentary on the interview process. Field notes may also comprise the preliminary thoughts the evaluator has about emerging patterns in the data.

Table 6. Advantages and disadvantages of observational and field notes

Advantages	Disadvantages
Taken at the time of observation or data collection which can aid memory recall	Inherent bias in notes comprising only what the evaluator sees and chooses to record
Providing data embedded in complex and dynamic contexts	Those being observed may change their behaviour in response to being observed (although this effect often diminishes with time)
May provide insights which are not contained in what people say they do, but direct observation of behaviour	Observation is time-intensive and requires careful planning and appropriate permissions
	Confidentiality or privacy concerns may limit evaluator access to some settings or events, including high level or politically sensitive meetings or clinical consultations

4.2.4 Open-ended survey questions

Although questionnaires typically consist of close-ended questions, they often include a small number of open-ended questions seeking more elaborate responses from participants. Open-ended questions allow participants to provide information beyond pre-defined response categories. There is some debate as to whether open-ended survey questions are a qualitative method. However, they are often used in evaluation contexts so have been included here.

Open-ended questions are asked when responses may be unsuitable for a close-ended format. For example, when the range is potentially wide; the topic, program or policy is new; or the answers are likely to involve multiple interacting factors. In formative evaluation, participants may be asked to name the most important health issue to address in their community. In process evaluation, participants may be asked what changes they would suggest to improve an initiative.

Table 7. Advantages and disadvantages of open-ended survey questions

Advantages	Disadvantages
Opportunity to get unanticipated responses across a large and perhaps representative sample of people	The data lack context, a critical element for interpreting data at the analysis stage
Can alert evaluators to subtleties in health decision making not possible through close-ended questions	The depth of the data are often limited, as many people do not want to spend the time to give a full response
Allows participants to convey their frustrations or wishes about an initiative	Those that are prepared to take the time may not be representative of the wider group
	Analysis can be resource intensive because the number of responses to be coded is large

4.3 Indigenist qualitative data collection methods

Box 5: Evaluation with Aboriginal people(s) and Indigenist research approaches

When conducting evaluations that involve Aboriginal people or report findings for Aboriginal people(s) as a group, it is necessary to ensure that culturally appropriate and safe processes are followed in accordance with Aboriginal Health and Medical Research Council of NSW requirements and other relevant guidelines (see Section 9). This includes ensuring that evaluations are led by or conducted in partnership with Aboriginal communities, that appropriate Aboriginal governance structures are in place, and that the evaluation results in a net benefit for Aboriginal people.[23] Evaluations of Aboriginal-focused initiatives should be conducted by Aboriginal evaluators wherever possible. Sufficient time, resources, and leadership by people with appropriate knowledge, skills, and experience are required to ensure good practice.

Indigenist research approaches respect and privilege Aboriginal ways of knowing, being, and doing, and aid in decolonising research practices.[26] Indigenist research approaches seek to redress power imbalances between participant and evaluator and actively centre Aboriginal people, cultures, and community control.[27,28] They also promote a strengths-based stance, recognising the capacities and capabilities of Aboriginal people.

For evaluations involving Aboriginal people(s), the appropriate use of Indigenist research approaches and Indigenist qualitative data collection methods can improve the experiences of Aboriginal participants, improve the quality of data collected, and improve the relevance and utility of evaluation findings. Wherever possible, qualitative data collection and interpretation should be conducted by Aboriginal evaluators or researchers. Aboriginal health evaluators and researchers are often highly skilled in community engagement, ensuring local cultural protocols are respected, Indigenist qualitative data collection methods, interpretation of findings, and supporting appropriate data sharing and sovereignty arrangements.[29]

Research Yarning and Dadirri are examples of Indigenist qualitative data collection methods.

4.3.1 Yarning

Yarning is a cultural process, not originally associated with research or evaluation, involving the telling of stories and sharing of knowledge through conversation. Research Yarning is an approach to single or group interviews which, while relaxed and interactive, is purposeful and ultimately aims to answer research or evaluation questions.[30] Compared to non-Indigenist interviewing methods, research Yarning involves more conversational interaction; that is, contributions from interviewers beyond asking questions and contributions from participants beyond responding to interviewer questions, are welcomed. This approach can be useful in building trust, exploring topics, and supporting reflection.

There are different types of Yarning which may be used in data collection. For example, rapport can be established through social yarning, and family yarning can establish personal connections of the participant and the interviewer (e.g. family, community, geographic, work, or sporting connections) to contextualise the knowledge shared in research yarns and establish understandings about the respectful treatment or sharing of that knowledge.[31,32] Yarning fits well with qualitative methods because of the emphasis on reflexivity, relationality, and co-constructed meaning between evaluator and participant. Beyond data collection, Yarning is foundational to relationship building and trust building essential in Aboriginal research and evaluation; Yarning supports genuine collaborations, information sharing, good governance and knowledge transfer.[31]

4.3.2 Dadirri

Dadirri is also a cultural practice not originally associated with research or evaluation. It is a term that refers to the Aboriginal practice of deep, respectful listening.[33] In research and evaluation contexts, Dadirri can refer to deep listening practiced during qualitative data collection, and it may also refer to an overarching approach that emphasises deep listening throughout an evaluation process, empowerment through the sharing of stories, and privileging of Aboriginal knowledges.

Figure 2. Implications for qualitative data collection methods in evaluation

Implications for engaging an independent evaluator

- Determine whether data already exists which may be provided to evaluators
- Ensure the evaluation timeframe allows for quality data collection and the team engaging the independent evaluator to review data collection processes and tools. Note that the time required and associated costs will depend on several factors such as data collection format (e.g. individual vs group), setting (e.g. online vs in-person), and respondent characteristics (e.g. where interviews are conducted in language and require translation and back-translation of transcripts)
- Ensure additional time is allowed for evaluations involving Aboriginal people(s) to ensure culturally appropriate and safe processes are followed
- Consider requesting inclusion of Appendix for technical details on data collection to ensure transparency

Implications for assessing proposals

Ensure the data collection methods and tools described:

- are appropriate to, and do not go beyond, the evaluation question(s) they are linked to
- are achievable within time and budget constraints
- can support the proposed analytic method
- are ethically and culturally appropriate

Implications for assessing reports

- Data collection methods should be described in sufficient detail for audience
- Discussion guides should balance appropriate level of structure with capturing unanticipated data
- Report should describe appropriate approvals for data collection (ethics, institutional, community)

5. Sampling in qualitative methods

Sampling refers to the recruitment of a portion of the group from whom evaluators would like to collect data. In quantitative methodologies, the main purpose of sampling is to maximise the ability to draw conclusions generalisable from the sample to the general population. In qualitative sampling approaches, the driving aim is data adequacy, which means sampling data sources with sufficient range and depth to adequately understand, explain or describe the topic of interest.[34] For example, the evaluation aim may be to describe the health experiences of one age cohort attending a health service. This description is not intended to be generalisable but has value in reflecting the lived experiences of service users in that age cohort, at that point in time.

Also, in contrast to quantitative methodologies, qualitative sampling is 'non-probabilistic', meaning the likelihood that someone will be selected is unrelated to their representation in the population, but rather their ability to provide required insights. For example, there are more women than men who use the NSW Get Healthy Information and Coaching Service, but men and women may be sampled equally for a qualitative study. In addition, it is acceptable for sampling in a qualitative study to be iterative rather than fixed at the start of a study, as the range of variation in the population is often unknown at the outset.

Sampling strategies for quantitative and qualitative studies also have some things in common; for both the choice of sampling strategy should maximise the generation of data that can answer evaluation questions while also considering the resources available for data collection.

5.1 Population selection

The first step in generating a sample is to select the population(s) from which the sample will be drawn. Population selection requires matching the evaluation questions to the groups who are best positioned to provide insights. Often in qualitative studies, this will mean selecting more than one population to understand the perspectives of multiple stakeholders. For example, we may want to know about a health service from the perspective of three distinct populations – service users, their families, and those responsible for service implementation.

It can also be the case that the 'ideal' population is unavailable, and a 'proxy' population may be selected who is not the ideal population but can provide insights in their place. For example, youth workers may be interviewed to gain insights on their clients' experiences of a policy, or families may talk about a family member's experience of a health service when the person is incapacitated. Proxy populations, however, should be selected carefully for being knowledgeable, but also acknowledged as giving accounts one step removed from the 'ideal' target population.

A range of sampling strategies are described below. While these strategies are described in terms of selecting people into qualitative evaluation studies, the principles can equally apply when sampling non-human data sources such as documents.

5.2 Purposive sampling strategies

The sampling techniques most commonly deployed in qualitative approaches are termed 'purposive sampling'. Selection is deliberate or purposeful, depending on considerations such as:

- whose perspective(s) are key to answering the evaluation question/s
- what subgroupings or characteristics are likely to affect people's experiences or opinions
- the accessibility of the target population (how willing and able they may be to participate in the evaluation)
- the overall size of the population.

The techniques described below are informed by one or more of these considerations. Note this list is not exhaustive and more than one technique may be used to sample within an evaluation.

5.2.1 Maximum variation sampling

Maximum variation sampling aims to get a range of opinions by sampling across groups and characteristics which may affect people's experiences or opinions. For example, sampling for interviews regarding people's experience of a statewide health service may ensure selection of young and old people, people of different genders, Aboriginal and non-Aboriginal people, people with and without

children, from English-speaking and diverse language backgrounds, from urban and regional areas and across socioeconomic strata. Such an approach allows the evaluation to capture what is common across different groups and what may be unique to people with certain characteristics by considering a range of voices.

5.2.2 Homogeneous group sampling

Homogenous group sampling, as the name suggests, seeks to narrow the characteristics upon which a group is selected. The aim is to explore in depth the experiences or opinions of a particular group with in-depth experience of a phenomenon, for example Aboriginal mothers' experiences of local maternal and child health services. This type of sampling is useful for constituting focus groups, where there is likely to be some commonality of experiences.

5.2.3 Whole population sampling

In circumstances where the population of interest is small, taking a sample may not be necessary and all members of a group are invited to take part. For example, there may be only a small number of people implementing a service, all with different roles, so it is appropriate to ask all to participate to yield a wholistic view.

5.2.4 Typical case sampling

Typical case sampling involves the selection of case/s who are typical of a population or organisation, or whose experiences are emblematic of the phenomenon under investigation. This means cases are not in any way atypical, extreme, deviant, or unusual. This type of sampling can be useful when there is a need to quickly identify and understand key aspects of a phenomenon as they manifest under ordinary circumstances. For instance, typical case profiles can help an evaluator gain a rapid understanding, or provide an in-depth insight, of how an initiative typically affects people's lives. Program staff and other key informants can assist to develop criteria of what is considered 'typical'.

5.2.5 Extreme case sampling

In contrast to typical case sampling, extreme or deviant case sampling aims to draw on experiences which comprise non-typical cases to make the causes of success or failure clearer. For example, cases may be selected with outstanding (positive) results from an initiative or where there has a been no or even negative impact (i.e. indicators have worsened since implementation). Their extreme nature may help identify clear mechanisms for success or failure.

5.2.6 Expert or key stakeholder sampling

Sampling people who have expertise in a particular phenomenon or who hold critical roles relevant to a topic is common in health initiative evaluation. Selecting experts or key stakeholders may be useful in formative evaluation in mapping the scope of a previously understudied issue. In outcome evaluation, an expert may give highly informative commentary on the performance of a health initiative because of the intimate role they played in its execution or because of their breadth of understanding of the topic.

5.2.7 Snowball sampling

Snowball sampling is often employed when the target population is difficult to identify for someone who is external or where potential participants may be reluctant to take part in an evaluation unless introduced by someone within their group. The 'snowballing' refers to the process where the sample grows through participants identifying and even contacting future potential participants rather than the evaluators having to identify all participants from the outset. For example, it may be difficult for an evaluator to identify all of the people who were involved in developing a particular health policy. The evaluator would rely on current study participants to identify others involved and iteratively invite them to be part of an evaluation. In another example, youths using illicit drugs may be unlikely to participate in the evaluation unless other members of their group or other trustworthy contacts can vouch for the experience and pass on the evaluator's details.

5.3 Other sampling strategies

Other sampling approaches prioritise practical considerations in sample selection over the characteristics of individual participants to a greater extent than purposive sampling strategies. However, population selection should still be consistent with the evaluation aims.

5.3.1 Convenience sampling

The term convenience sampling refers to recruitment which occurs through channels which are convenient to the evaluator. For example, a physical activity promotion app designed for young adults may be piloted with first year university students because they are straightforward to recruit for a university-based evaluator. The population in this example is still

appropriate, as it includes those who are among the target population. However, convenience samples can lead to a narrow range of reported experiences when a diverse range may be required to properly address the evaluation question.[14]

5.3.2 Opportunistic sampling

Opportunistic sampling is where sampling takes place at an event or under circumstances where the population of interest becomes easier to identify or more accessible than usual. Opportunistic sampling has some cross-over with convenience sampling in that it has a practical element but is more strategic because it is designed to exactly capture the target population. For example, it may be difficult to identify or recruit a diverse sample of cyclists to evaluate the impact of minimum passing distance laws on cycling habits. Attending an event such as a mass participation cycling event (e.g. Bicycle NSW Spring Cycle) to recruit participants is a targeted and efficient way to recruit a diverse group with relevant characteristics.

5.3.3 Quota sampling

Quota sampling aims to ensure the sample includes certain groups who may be underrepresented if a convenience sample were taken, or even maximum variation sampling. The 'quotas' are decided prior to data collection, with the assumption that the quota nominated will yield adequate data. For example, an evaluation of a falls prevention program may have quotas for sampling residents in retirement villages, low, or high-level residential care facilities, with a third of the final sample coming from each group.

5.4 Sample size

Determining the appropriate sample size is less straightforward for qualitative methods than for quantitative methods. Notably, sample size for qualitative methods should not be measured against quantitative notions of obtaining a representative sample or statistical power. Rather, as described previously, the driving aim for qualitative sampling is to achieve data adequacy. Therefore, sample adequacy rather than size is of foremost importance; sample adequacy may be achieved with a relatively small sample size. In qualitative evaluation, face validity is also a relevant consideration. This refers to whether a sample size appears adequate for findings to be perceived as credible (i.e. does the sample 'look like' it can provide sufficient data to answer the evaluation question).

For some sampling approaches (e.g. maximum variation sampling), sampling adequacy is associated with the notion of saturation; popularly conceived as the point at which 'no new information' is being generated and therefore further recruitment is considered redundant.[35] The point at which saturation is reached depends on the detail and volume of data available and the variability of the sample. This is often unknown at the start of a study so the final sample size may be determined iteratively as data is collected and analysed. For other sampling approaches (e.g. whole of population, extreme case, typical case, quota sampling) saturation is not relevant as data adequacy (and subsequently sample size) is predetermined by the sampling approach.

From a practical point of view, not knowing the final sample size required for sample adequacy at the start of an evaluation can have implications for planning and budgeting. Having clear evaluation questions and clearly identifying groups of particular interest should help choose the right sampling strategy (i.e. maximum variation, extreme case, whole population, etc.) and guide a more accurate assessment of likely sample size and associated costs. For example, if a maximum variation sampling approach is chosen, then considering the number of stakeholder groups that will be engaged (e.g. managers and operational staff, consumers and other stakeholders impacted by the initiative, such as carers or peak bodies) and the expected variation of views within each group (e.g. experiences amongst service users) will help determine when sample adequacy may be reached and subsequently sample size. Whereas narrowly focused evaluation questions, best addressed by a specific group (e.g. influential stakeholders), suggest a smaller sample size. Flexibility in evaluation budgets may be required to accommodate changes and staged estimates involving increasingly comprehensive data collection scenarios can be useful. It may be helpful to seek advice from internal evaluation advisors about appropriate resourcing for specific circumstances.

Ultimately, sample size should be a result of the coverage required by the evaluation question(s) and the purpose to which the findings will be put. There should also be sufficient description of the methods and the resultant sample to allow for an assessment of the likelihood that reported conclusions are derived from sources with adequate range and depth.

Figure 3. Implications for qualitative sampling in evaluation

Implications for engaging an independent evaluator

- Consider the need for policy makers or service providers to facilitate access to key populations (key stakeholders, program participants), and plan resources accordingly
- Ensure sufficient budget and time given for adequate and possibly iterative sampling
- Specify any sampling requirements (e.g. inclusion of priority populations)

Implications for assessing proposals

Ensure the sampling approach described:

- reflects the appropriate level of narrowness or broadness needed
- balances time/resource constraints with sample adequacy
- addresses the diversity of factors affecting opinions
- is realistic about accessibility of target population(s)
- addresses iterative sampling if timeline allows
- clearly explains how data saturation will be tested and reached, or why it is not relevant (e.g. whole population sample)

Implications for assessing reports

- Sampling approach and outcomes clearly described
- Limitations of sampling are acknowledged
- Saturation clearly described if used to justify ultimate sample size
- Description of sample on key characteristics provided

6. Analysis of qualitative data

Analytic methods for qualitative data cover a spectrum from the more descriptive, categorisation of data to highly interpretive development of theory.[36] For descriptive analysis, data are largely taken on face value, while for interpretive analysis, potential underlying meanings and messages are examined. Some methods draw on existing theory to structure analysis (deductive) and others aim to generate theories or explanations "grounded" in the data collected for a particular study (inductive). In many cases, a combined approach is used, when some issues are of known interest before analysis begins, but space is also left to discover other unexpected findings during the analysis process. The type of analysis and where it may fall within these spectrums (deductive and inductive, descriptive and interpretive) will depend on the evaluation questions, the purpose to which the evaluation findings will be put, and the discipline (e.g. psychology, sociology, feminist, Indigenist) from which the analytic techniques are drawn.

6.1 Preliminary analysis

Analysis of qualitative data is often described as 'iterative' as the evaluator generates, tests and revises their analysis with ongoing exposure to the data and the insights of others in the evaluation team. Often the analytic process commences while data are being collected as the evaluator listens to their interviewees and tries to integrate what they hear with their prior conceptions and previous interviews. As the evaluator re-listens to recordings, transcribes and/or corrects transcriptions, the preliminary analysis continues as a process commonly called 'data immersion'. Immersion in the raw data improves understanding and is critical to a high-quality analysis, regardless of the analysis approach adopted. This may require reading documents, notes or transcripts, or listening to audio recordings multiple times, which can be a time consuming and resource intensive process. Preliminary analysis enables the evaluator to become familiar with whole dataset but is not a substitute for systematic and formal analysis.

6.2 Formal qualitative analysis

Once the data are prepared (e.g. interview transcripts), and the preliminary, more informal analysis is complete, the formal process of analysis begins. One way to think about the process of qualitative data analysis is that it deconstructs then reconstructs the raw data to account for both commonalities and exceptions across the dataset. Two important differences to quantitative analysis emerge from this conceptualisation. First, the analysis does not replicate the structure of data collection (i.e. the questions) but reorganises the data across questions and participants. Before qualitative data analysis software existed (see Box 7), an analyst might literally cut up printed transcripts into chunks with a pair of scissors and rearrange them. Second, exceptions are not considered 'outliers'. In contrast to quantitative analysis, the frequency with which something is mentioned is de-emphasised in qualitative approaches in favour of explanatory power – one person's observation may be key to understanding a phenomenon.

It is beyond the scope of this guide to describe the large range of qualitative analysis techniques currently used. This section will focus on two techniques which are frequently used in evaluation, namely thematic and content analysis. Other analytic approaches used in evaluating health initiatives include 'narrative analysis' and 'framework analysis'. Narrative analysis examines complete participant stories about their lived experiences. Framework analysis is a structured approach using a grid to organise data by theme and case, which may be useful when a site (e.g. hospital) or the characteristics of individuals are informative in interpreting the analysis. Further details about these and other techniques may be found in Liamputtong (2019).[1]

6.2.1 Thematic analysis

Thematic analysis is often used in the evaluation of health initiatives and has been defined as "a method for identifying themes and patterns of meaning across a dataset in relation to a research question" (Clarke and Braun (2013),[36] p175). Thematic analysis is a flexible yet systematic approach accessible to evaluators across a range of skill levels.

The six-step process of thematic analysis as described by Braun and Clark[37] is shown in Figure 4. Rather than linear, the steps are intended to be iterative where the evaluator goes through the process of refining the codes and themes using multiple passes through the data, checking with other evaluators in the team and participants. Reports and proposals should clearly describe the process that is followed to allow users of an evaluation to decide whether the conclusions drawn are reasonable.

Although this guide does not aim to specify 'how to' do thematic analysis, two key steps (coding and theme development) that comprise the heart of thematic analysis are described in more detail below.

Figure 4. A six-step process for thematic analysis

Adapted from Braun and Clark[38]

6.2.1.1 Coding

After preliminary data analysis has been conducted (see 6.1 and Step 1, Figure 4), an evaluator moves to the 'coding' process (Step 2, Figure 4). Coding is a systematic process that involves labelling pieces of the data (transcripts, documents) in terms of what that piece of text might mean. For example, an evaluation with participants in a weight loss maintenance program might label pieces of interview text (the raw data) with codes such as 'participant feelings about weight gain' or 'handling of special occasions'. The pieces of text being coded may be small (a fragment of a sentence) or large (a whole paragraph) and some pieces may have more than one code. Each code can usually be applied to multiple pieces of text (or multiple parts of the data). Some codes can be used to tag pieces of text which may be retrieved for the purpose of describing the sample, such as people's roles, rather than relating to the thematic analysis.

Codes can be developed iteratively, usually on a subset of the data (inductive analysis). Finalised codes are integrated together into a coding framework which is then applied across the full dataset. Alternatively, a coding framework may be chosen prior to analysis (deductive analysis).

Coding (and theme development, see 6.2.1.3) may be conducted by a single person or multiple people depending on the size of the dataset and the resources available. Where possible, having multiple evaluators collaborate when developing and applying codes and themes can strengthen the analysis process and ensure sense checking and selection of the themes which best account for the data in relation to the evaluation question. Such collaborative processes are often only done on a sub-set of the data, not the whole dataset. Reporting should describe the process of how coding differences are managed.

Box 6: Inductive versus deductive analysis

A key concept in qualitative analysis of many kinds is whether the analysis will be inductive, deductive, or a combination of both. Inductive analysis refers to an analytic process where codes are generated from the data rather than starting with a prespecified framework or theory. Codes are not set before commencing the analysis but developed as part of the analytic process. It can be thought of as a 'bottom up' coding approach.

Deductive approaches involve coding data using pre-defined concepts derived from a theory or framework. It can be thought of as a 'top down' coding approach. This can be a useful technique for a range of purposes, for example to test whether a theory of behaviour change, which has underpinned a health program's development, is demonstrated in participants' accounts of their experiences of a program. In other cases, it may be used where a program or service has been modified to meet principles of, for example, cultural inclusion. A deductive analysis approach could examine whether those principles are present in observations of the program implementation, program protocols and/or interviews with participants and program managers.

In any analysis, it is likely that an evaluator will use both deductive and inductive processes. At the very least, in an inductive analysis, an evaluator narrows their focus to the evaluation question at hand and brings their prior knowledge or experience. Similarly, a deductive analysis may start with a coding framework, whether inspired by theory or some other set of principles, but other codes will also likely be developed inductively and incorporated into the interpretation of the data. Therefore, any one analysis could be located on a spectrum depending on the dominance of inductive and deductive methods.

6.2.1.2 Theme development

Theme development builds on the coding process and is a more interpretive step. Themes are often abstract entities or ideas, capturing implicit concepts 'beneath the surface' of the data, but can also capture more explicit and concrete meaning.[38] Potential themes are tested for their applicability across the dataset and how well they provide insights to answer the evaluation question.

Themes differ from codes. Codes refer to one unit of meaning, whereas themes bring together a diversity of codes which speak to the same unit of meaning. For example, an evaluation of the NSW Quitline had 'professionalism' [of the service] as one theme for understanding how clients gauged their satisfaction with Quitline.[39] To the clients, professionalism was reflected in the quality of the information they received and whether staff adhered to expectations of service standards (two subthemes of the professionalism theme). The theme brought together codes across the dataset marking the clients' experiences and reflections (positive and negative) of different components of the service (e.g. advice, printed information, call back protocols) which shared the meaning of professionalism as it was conceived by the clients.

6.2.1.3 Themes or topic summaries?

A common occurrence in qualitative analysis is the development of 'topic summaries' when a thematic analysis was intended. Topic summaries are derived from a descriptive analysis of text/s on face value; that is, a summary of what was reported/recorded. Themes, on the other hand, are interpretative and meaning-based and could not be developed prior to data collection based on assumptions or theory.[40] Accordingly, theme names are meaning-oriented such as 'validation of personhood' or 'proving gender'.

If a 'theme' maps closely to a data collection question, or could have been developed prior to analysing the data, then it is likely to instead be a topic summary.[40] Another clue that a theme may actually be a topic summary is when it consists of one or two words, like: "Doctors", or "Experiences of…", "Barriers to…", "Influences on…"; suggesting that diverse experiences, barriers and influences have been grouped under this heading.[40]

The development of either themes or topic summaries may be appropriate, depending on the evaluation questions that require answering. Topic summaries and other less interpretive analytic techniques can at times be fit for the purpose of answering more descriptive evaluation questions such as documenting how an initiative is implemented at different sites,

noting what adaptations are made, or the key features of a model of care. For evaluations where mechanisms, drivers and explanation are desired, topic summaries generally do not capitalise on two of the main advantages of qualitative data: depth, and capacity to yield new insights into a phenomenon. Results which comprise a series of topic summaries seldom tell an integrated story of key mechanisms underlying a phenomenon, or higher-level principles which can inform future initiatives beyond the current one under investigation, as a thematic analysis can do.[38] In either case, evaluators should understand this difference, accurately describe their methods, and not claim that thematic analysis has been conducted if a topic summary is presented.

6.2.2 Content analysis

Content analysis is a technique often used in analysis of communication, where the primary focus is on the language used. In the health initiative context, it may be applied to interview data, but it is used more often in analysis of media articles, comments on social media, advertising, website content, and policy documents regarding health. For example, content analysis has been used to analyse compliance with regulations for internet advertising of infant feeding products.[41] Another study has examined media reporting on tobacco plain packaging over a six-year period and content analysed commentary in terms of whether the articles were supportive, opposing, neutral or mixed.[42]

Content analysis could be seen as based on either words or codes as the unit of analysis. The coding approaches in content analysis use the whole text, retaining whatever context is given to generate codes. As with thematic analysis, coding may use a framework developed from the data (inductive) or beforehand (deductive). For example, the infant-feeding formula study above used an existing food labelling standard to develop their coding frame.[41]

Content analysis, while still focusing on words as the unit of analysis, often involves a more quantitatively oriented approach to identify and count the use of certain words or co-occurrence of words. The analytic process can be semi-automated in that software can detect and report the frequency of certain expressions or words, as well as the proximity of words (e.g. Prior 2014).[43] Data may also be presented as proportions, as in the plain packaging example. Care should be taken, however, to ensure presented data is valid and meaningful by appropriate sampling and rigorous approaches to coding.

6.2.3 Analysis of open-ended survey questions

Open-ended survey questions, as described in Section 4, may be used to collect qualitative data across a survey sample to capture greater detail and insight on an evaluation topic. There are, however, limits to the value of these data compared to longer form qualitative data collections. Which analytic technique is suitable will depend on the richness of the data and the evaluation question.

Box 7: Qualitative data analysis and technology

Technology in qualitative research and evaluation can assist in efficient data generation and management. Video-conferencing technology, especially during and since the COVID-19 pandemic, has become a common means by which interviews and even focus groups are conducted. It allows people to participate who may not have been able to due to geographic or time constraints. Many platforms also have the capacity to generate real-time transcripts, although as with any transcripts they still need careful checking for accuracy.

Qualitative Data Analysis Software (QDAS) is another technological advancement in qualitative data analysis, albeit less recent. The most widely used is NVivo which assists evaluators to manage and analyse a range of data sources including transcripts, journal articles and existing documents (e.g. policy and protocol documents). QDAS does not 'do' the analysis any more than a statistics package 'does' a quantitative analysis. Both systems require a skilled and knowledgeable evaluator who uses the software to facilitate analysis. At its most basic level, QDAS simplifies data coding and retrieval processes.

Generative Artificial Intelligence (AI) is starting to impact the knowledge generation sector and is being used in some studies to 'scrape' large amounts of data from publicly available sources (e.g. social media platforms) and at times to conduct rudimentary analysis of this data. However, the utility of AI for qualitative analysis – and for research generally – is somewhat constrained by its limited capacity to use judgement and draw on context.

It is also important to note that submitting confidential transcripts to public platforms is unethical and new technologies should only be used for qualitative analysis if appropriate privacy and security measures are in place.

Thematic analysis, although possible, may be unsuitable in many cases because of the lack of contextual information available in such data.[44] Content analysis may be more suitable for sparse data and where the evaluation question concerns communication.

Figure 5. Implications for qualitative analysis in evaluation

Implications for engaging an independent evaluator

- Specify if particular analytic techniques are preferred
- Provide sufficient time and resources to support sound analysis

Implications for assessing proposals

The analytic approach should:

- be specified and appropriate to the data collected
- be consistent with any overarching theoretical approach
- be achievable in the timeframe
- have sufficient resources (technical and human) allocated
- be appropriately inductive or deductive
- be carried out by a team with sufficient expertise and experience in the analysis techniques described
- be appropriate to the evaluation question

Implications for assessing reports

- The description of the analysis should contain sufficient detail to assess rigor and appropriateness
- Analytic processes should be systematic and ideally collaborative for validation purposes
- Results should be logically structured and integrated
- Exceptions should be described and integrated into analysis
- Conclusions should be consistent with content and the level of analysis

7. Reporting of qualitative findings

This section covers some of the key issues associated with presenting qualitative methods and findings in evaluation reports. The checklist in Section 8 outlines features that should be present in reports that contain qualitative evaluations. The NSW Ministry of Health's *Preparing and appraising evaluation reports: A checklist* also includes points on this topic and can be used in conjunction with the information in this guide to ensure evaluation reports for NSW Health initiatives are of high quality and relevant to decision making. As with all evaluation reporting, the reporting of qualitative studies should reflect a consistent and convincing path from the questions asked, methods used, through data presentation to interpretation and recommendations.

7.1 Description of context and methods

While ensuring the rigour of any evaluation remains a priority, most real-world evaluations will unavoidably contain constraints that affect the comprehensiveness and validity of findings and conclusions presented. To enhance the usefulness of evaluation findings, it is important that evaluation reports contain clear and detailed descriptions of the methods used – including descriptions of sampling, data collection and analysis processes – and any associated limitations. This remains true for both qualitative and quantitative evaluations. Clear and comprehensive reporting allows the reader to interpret the relative strength and usefulness of the findings and make informed decisions about how they may be applied. It is also important that adequate information is reported about the context within which the health initiative was implemented and evaluated. Again, this helps to clarify the meaning and relevance of the reported evaluation findings and can provide insights about the transferability of the findings to other practice contexts and settings.

7.2 Presenting evaluation findings

The way in which evaluation findings are presented will depend on factors such as the methods employed, the questions to be addressed and the audience. An overview of some of the key considerations for reporting on evaluations that either employ only qualitative methods, or a mixed methods approach, is provided below. Overall, the findings should be presented so the reader can easily understand the answer to the evaluation question(s).

7.2.1 Presenting qualitative evaluation findings

Thematic analysis is frequently used in qualitative evaluations of health initiatives. Findings from a thematic analysis are usually presented with themes as headings and sub-themes as subheadings with explanations of what they mean and how they link together (findings should not be presented as an unintegrated list of themes). Whether a true thematic analysis or more superficial topic summary is provided should be considered (see Section 6).

Some reports will include findings (themes) and discussion of findings, in regard to their implications for the initiative, in separate sections. Other reports may integrate the two by including commentary on the significance of the themes as they are being explained. This may be followed by a short conclusion on major implications and recommendations for practice change in line with the purpose of the evaluation. Both are acceptable approaches. It may also be appropriate to describe the findings in relation to what is already known about the concepts raised, as would be done in an academic journal article. Regardless of the structure used, the presented findings should be credible and compelling with clear relationships evident between the data and conclusions that have been drawn.

A summary of viewpoints is often provided when describing themes. However, it is not appropriate to state the proportion of people who said something; participants are not sampled appropriately for this purpose and the data collection process is fluid rather than standard between interviews or focus groups. Instead, the text may note if "many" or "some" people raised a concept to convey how common a viewpoint was. Outlying or contradictory views should also be reported alongside those of the majority, noting that popularity does not signify importance; outlying views can also be illuminating. As qualitative methods focus on eliciting experiences and perceptions, it is important that findings are framed appropriately (e.g. "participants reported" or "participants felt") and that viewpoints are not presented as fact.

Quotes also play an important role and should be used not only to illustrate themes but to allow the reader to assess the reasonableness of the interpretation. Quotes should be used to support or illuminate the evaluator's analysis, and should not form the bulk of a results section. Quotes should be de-identified using pseudonyms or labels (such as "clinician R4") and be drawn from a variety of individuals/documents across the sample throughout the report. Care should also be taken to ensure quotes do not contain other information that may inadvertently identify a participant.

7.2.2 Presenting mixed method evaluation findings

Where mixed methods are employed, good reporting practices for qualitative and quantitative studies still hold. However, the report should also describe the sequencing and integration of qualitative and quantitative components and integrate qualitative and quantitative findings in a useful way. How this occurs depends on the purpose of using a mixed methods approach (see Section 3.4) and the stage at which findings are integrated. Most commonly in mixed method evaluation, integration of findings occurs at the stage of interpretation and conclusion rather than data collection or analysis.

Qualitative and quantitative findings can be brought together on a topic-by-topic basis or presented in different sections, or some combination of the two.[21] For example, the results for different components of the evaluation (e.g. administrative data collection, surveys and interviews) could be presented separately and then brought together by general topics or themes, to address key evaluation questions, in the discussion or key takeaway sections. Alternatively, qualitative and quantitative findings can be presented together by theme, concept or evaluation question throughout the report. This approach may be used if qualitative methods have been employed to explain or enrich findings of a quantitative analysis or vice versa. In cases where triangulation of methods is the goal, the findings from the analysis of each dataset should be compared to determine if there is a convergence of findings.[45] Divergent findings should also be reported. They may help to clarify the results of an evaluation, or it may be necessary to make a judgement to weigh the relative credibility of the data sources and the quality of the data to resolve discrepancies.[46]

Regardless of how findings are integrated, it is important that evaluators have sufficient time to appropriately analyse the datasets from each method employed so they can be adequately reported. Findings from each method used to collect data should be reported, or methodological justification provided for why this has not occurred. Care should also be taken not to over or under-utilise a particular dataset; for example, where quantitative data is more easily reported than qualitative data. It should also be clear how each dataset has contributed to explanations and conclusions.

8. Checklist

8.1 Introduction

The checklist for qualitative methods in health initiative evaluation below can assist in assessing the completeness and appropriateness of evaluation proposals and reports that use qualitative methods. While it draws on existing checklists for qualitative research such as the Consolidated Criteria for Reporting Qualitative Research (COREQ)[47] it has been adapted for policy/practitioner application to health initiative evaluations rather than for academic purposes. The checklist should be used in conjunction with a broader review of the quality and appropriateness of the evaluation, and compliance with the objectives and requirements of any tender documents. For assessing evaluation reports specifically, readers should also refer to the NSW Ministry of Health's *Preparing and Appraising Evaluation Reports: A Checklist*.[4]

8.1 Using the checklist

The first column in the checklist describes a component of the qualitative methods or procedure relating to the use of qualitative methods phrased as statements, under subheadings. The next column provides for a rating as to whether that component of the methodology or approach was described, and the final column asks for an assessment of the appropriateness or quality of the component. Not all items will be relevant to all evaluations; some may be relevant for proposals and reports, and others for proposals or reports alone. Many of the items relevant to proposals will also be relevant to evaluation plans if a proposal is not received. Although the columns ask for rating items, the checklist is not designed to generate a score. The rating scale allows that proposals and reports may provide some information on an item but it may not be complete. Where there is missing or insufficient information, this should serve as a prompt to the team managing the evaluation to request further information or, in the case of proposals, signal an unsuitable bid.

8.2 Checklist for qualitative methods in health program evaluation

Key

Rating: Y = Yes (fully); N = No (not at all); P = Partial; na = not applicable

Pr/R = suitable for assessing proposal and/or report

Note

The items for proposals may also serve as a checklist for an evaluation plan once independent evaluator has been engaged. At that stage, there should also be an assessment of data collection tools and protocols, documents which are usually not available at the proposal stage.

Item	Clear responses	Information provided Y/N/P/na		Appropriate Y/N/P/na	Comment
Evaluators					
Evidence of experience and training of the evaluation team in the use of qualitative methods in evaluation provided (Pr)		na		na	
Ethical considerations					
Requirements for ethical review addressed or discussion of consideration of ethical conduct if no ethics review required (Pr)		na		na	
Approval of authorising research ethics committee noted, if required, or reason specified if ethics not required (R)		na		na	
Methods: Approach					
Use of a qualitative approach is justified, including in mixed methods (Pr)		na		na	
Sufficient resources (technical and human) allocated and approach feasible in the timeframe (Pr)		na		na	
[For mixed methods evaluation only] Sequencing and integration of qualitative and quantitative components described (Pr) (R)		na		na	
Methods account for the needs of specific populations included in the evaluation (Pr) (R)		na		na	
Methods: Sampling/participant selection					
Population/s from which participants are selected is described or, in the case where documents are the data, the population of documents (e.g. policies, media articles, websites) described (Pr) (R)		na		na	
Sampling strategy described (e.g. maximum, convenience, snowball) (Pr) (R)		na		na	
Methods: Data collection					
Data collection methods (e.g. interviews, participant observation) and rationale for their use described (Pr) (R)		na		na	
Data collection procedure described (e.g. interview guide development, piloting, consent process, who collected data, how (face-to-face/telephone/video call)) (Pr) (R)		na		na	

Item	Information provided Y/N/P/na	Appropriate Y/N/P/na	Comment
Data collection tools provided (e.g. interview questions, prompts, discussion guides, data extraction protocols) (R)	na	na	
Interviews and focus groups audio or video recorded and transcribed verbatim, where used (Pr) (R)	na	na	
Methods: Data analysis			
Data management and analysis processes (e.g. double coding) described, including any software system used and personnel involved (Pr) (R)	na	na	
Analytic approach (e.g. thematic, content, inductive or deductive) and rationale described (Pr) (R)	na	na	
How findings have been synthesised described (e.g. description of principles and choices informing the formation of patterns and categories; how were major and minor themes developed) (R)	na	na	
Data saturation discussed and defined, where relevant (Pr) (R)	na	na	
How interpretation was assessed for credibility and confirmability described (e.g. member checking, sense checking/ collaboration in analysis and triangulation) (R)	na	na	
Reporting			
Total sample size and non-participation rate reported (R)	na	na	
High-level participant characteristics reported (e.g. number of participants from each site/in each category) (R)	na	na	
Participant quotations identified by participant number and presented to illustrate the themes/findings (R)	na	na	
Range of voices and views (including dissenting views) represented in quotes where applicable (R)	na	na	
Findings presented in a way that addresses the evaluation objectives (R)	na	na	
Analysis presented is credible and compelling (i.e. themes flow logically from the findings; relations between data theory are described; interpretations are insightful) (R)	na	na	
Results/discussion considers contradictory or diverse cases (R)	na	na	
Results/discussion considers major and minor themes (R)	na	na	
Results/discussion explore corroborative findings (e.g. triangulation) (R)	na	na	
Findings contextualised with an understanding of the initiative being evaluated and evaluation purpose (R)	na	na	
Implications and recommendations justified based on data presented (R)	na	na	
Evaluation strengths and weaknesses identified with sufficient information to enable credibility to be assessed (R)	na	na	

Items adapted from Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007; 19(6): 349-57; A reporting guide for qualitative studies. *Can Comm Dis Rep* 2016; 42: 177-8. https://doi.org/10.14745/ccdr.v42i09a02; and Fossey E, Harvey C, McDermott F, Davidson L. Understanding and evaluating qualitative research. *Aust N Z J Psychiatry* 2002; 36(6): 717-32. Others were developed specifically for this guide.

9. Key resources and further reading

For further information and broader guidance about the planning, conduct, reporting, and use of evaluations, please see the resources listed below.

Name of resource	Brief description
CEE Guides	
Planning and Managing Program Evaluations: A Guide	This guide promotes a proactive and structured approach to planning evaluations, including information on when and how to engage an independent evaluator.
Developing and Using Program Logic: A Guide	This guide promotes a planned and structured approach to developing program logic and includes information on: the meaning and purpose of program logic, when and how to develop program logic, and how program logic can be used, with a particular focus on planning an evaluation.
Study Design for Evaluating Population Health and Health Services Interventions: A Guide	This guide supports NSW Health staff in the planning of evaluations of interventions using appropriate study designs with a focus on quantitative study designs.
Engaging an Independent Evaluator for Economic Evaluations: A Guide	This guide has been developed to support NSW Health staff engage an independent evaluator for economic evaluations of health programs, particularly those in population health. The guide should be read in conjunction with Planning and Managing Program Evaluations: A Guide.
Planning Economic Evaluations: A Checklist	The purpose of this checklist is to assist users to systematically review the quality and relevance of economic evaluations. The checklist focuses on the core principles of economic evaluation and how each can be used in appraising economic methods found in a range of documents such as peer-reviewed journal articles, grey literature, project proposals and reports.
Preparing and Appraising Evaluation Reports: A Checklist	This checklist promotes a rigorous and planned approach to the preparation and appraisal of evaluation reports. It includes criteria for ensuring that reports are complete, that the results are robust and the conclusions are sound.
Treasury Guides	
NSW Treasury Policy and Guidelines: Evaluation TPG22-22	This document sets out mandatory requirements, recommendations and guidance for NSW General Government Sector agencies when planning and conducting evaluations.
NSW Treasury Evaluation Workbooks 1–8 and templates	These workbooks contain information to support monitoring and evaluation including templates for program logic models, data matrices, project management, report planning and reporting. The workbooks are: Workbook 1: Foundations of evaluation Workbook 2: Monitoring and evaluation framework Workbook 3: Evaluation plan: Design the evaluation Workbook 4: Evaluation plan: Manage the evaluation Workbook 5: Evaluation plan: Use the right expertise Workbook 6: Evaluation plan: Report and use findings Workbook 7: Example evaluation report template Workbook 8: Complex initiatives Workbook templates

Name of resource	Brief description
NSW Treasury Evaluation technical notes and glossary of terms	NSW Treasury have produced detailed technical notes and guidance on the terms and methods used in evaluation including. Resource – Glossary of terms Technical note – Sampling strategy Technical note – Outcome evaluation design Technical note – Evidence in evaluation Technical note – Ex-post cost benefit analysis
Aboriginal research and evaluation ethical guidelines and appraisal tools	
Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders; National Health and Medical Research Council	These guidelines provide a set of principles to ensure research is safe, respectful, responsible, high quality and of benefit to Aboriginal and Torres Strait Islander people and communities.
Keeping research on track II: a companion document to Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders; National Health and Medical Research Council	This guide was developed to provide advice on how the values and principles outlined in Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders can be put into practice in research.
NSW Aboriginal Health Ethics Guidelines: Key Principles (2023); AH&MRC Ethics Committee	The purpose of this document is to ensure that research that affects Aboriginal people and communities is done in a culturally appropriate way, involves and considers the people that it affects. This guideline may be useful for researchers that are planning to submit a research application for review by the Aboriginal Health and Medical Research Council (AH&MRC) Ethics Committee.
AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research	The Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) Code of Ethics outlines four principles that underpin ethical Australian Indigenous research: Indigenous self-determination, Indigenous leadership, impact and value, and sustainability and accountability.
Indigenous Evaluation Strategy and companion *Guide to Evaluation;* Australian Government Productivity Commission	The *Indigenous Evaluation Strategy* provides a whole-of-government framework for Australian Government agencies to use when selecting, planning, conducting, and using evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people. The accompanying *Guide to Evaluation* provides practical advice for Australian Government agencies on how to implement the Strategy when conducting evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people.
Consolidated criteria for strengthening reporting of health research involving indigenous peoples: the CONSIDER statement	The CONSIDER statement provides a checklist for the reporting of health research involving Indigenous peoples to strengthen research praxis and advance Indigenous health outcomes. Eight domains for reporting research involving Indigenous Peoples are discussed in this paper: (i) governance; (ii) relationships; (iii) prioritization; (iv) methodologies; (v) participation; (vi) capacity; (vii) analysis and findings; and (viii) dissemination.

Name of resource	Brief description
The Aboriginal and Torres Strait Islander Quality Appraisal Tool	The Aboriginal and Torres Strait Islander QAT appraises research quality from the perspective of Aboriginal peoples. The tool covers several aspects of evaluation: setting appropriate research questions; community engagement and consultation; research leadership and governance; community protocols; intellectual and cultural property rights; the collection and management of research material; Indigenous research paradigms; a strength-based approach to research; the translation of findings into policy and practice; benefits to participants and communities involved; and capacity strengthening and two-way learning.
Ethics Guidance	
National Statement on Ethical Conduct in Human Research 2007 (updated 2018); National Health and Medical Research Council	The National Statement sets national standards for use by any individual, institution or organisation conducting research with human participants.
Getting ethics approval; Justice Health and Forensic Mental Health Network	Guidance on when human research ethics committee review is required for evaluations.
Low and Negligible Risk Research Guideline GL2023_007; NSW Health	This Guideline represents NSW Health's interpretation of the National Statement on Ethical Conduct in Human Research (2007) (National Statement) as it applies to low and negligible risk (LNR) research. It has been developed to clarify the requirements of LNR research under the National Statement. Section 7.1 provided a decision tree for LNR review.
Ethical Considerations in Quality Assurance and Evaluation Activities 2014; National Health and Medical Research Council	Quality assurance (QA) and evaluation are important to ensure effective work and the best outcomes. However, confusion arises over if an activity is research, evaluation or QA as there may be similar research methods used. This document assists organisations in developing QA policy and appropriate oversight.

10. Key definitions

Bias in research arises from deviations from the truth during any part of the research process (e.g. data collection, data analysis, interpretation or publication), causing distorted results and potentially false conclusions to be drawn.[9]

Close-ended questions refer to questions where participants choose from a distinct set of pre-defined responses, such as 'yes/no' or a limited range of likely answers.

Coding frameworks arrange codes in relation to one another (e.g. a hierarchy with sub codes under other broader codes) and provides very brief definitions of codes.

Confirmability in qualitative research refers to whether findings and interpretations reflect the views of participants.[14]

Credibility in qualitative research refers to comprehensive, trustworthy and defensible explanations of the data.[14]

Dadirri (da-did-ee) is a word, concept, and spiritual practice from the Ngan'gikurunggurr and Ngen'giwumirri languages of the Aboriginal peoples of the Daly River region of the Northern Territory, Australia. All Australian First Nations have their own word for deep listening, meditation, knowing, and reflecting. For example, in the Wiradjuri language the word is Winhangadhurinya.[48]

Data adequacy is achieved by sampling data sources with sufficient range and depth to adequately understand, explain or describe the topic of interest.[34]

Datasets are the complete set of transcripts or documents comprising the data collected.

Decolonising approaches recognise that the way of knowing has been historically and institutionally contrived in a Western construct, and that Indigenist approaches, methodologies and methods can be used to shift the research paradigm and privilege the knowledge and experiences of Indigenous peoples.[32,49]

Dependability in qualitative research refers to the coherence between methods and findings, and transparency and auditable research process.[14]

Ethical considerations are a set of principles that guide research designs and practices. These principles include voluntary participation, informed consent, anonymity, confidentiality, potential for harm, and results communication.[22]

Evaluable refers to whether something can be evaluated.

Generalisability in research refers to whether the results can be applied to a broader context or population than the one studied. Generalisability is determined by how representative the sample is of the target population.

Indigenist research approaches respect and privilege Aboriginal ways of knowing, being and doing, and aid in decolonising research practices. Indigenist research approaches seek to redress power imbalances between participant and evaluator and actively centre Aboriginal people, cultures, and community control.

Interview schedules are sets of questions to serve as a guide for the interviewer to cover the topics relevant to the evaluation question.

Member checking refers to the process whereby participants are asked to check transcripts and/or interpretation of data prior to finalisation. There are also 'member reflections' which can involve sending participants the raw data and/or researcher interpretations for collaboration and feedback. It allows participants to revise their opinions as well as have input into how their words are interpreted and contribute meaningfully to research findings.[50]

Mixed method evaluation refers to the use of qualitative methods in combination with quantitative methods.

Observational notes are structured data collection tools which usually have pre-defined and observable categories established prior to the study commencing. In contrast, during unstructured observation, researchers seek to capture everything that occurs within the setting without any predefined categories. This method is often used to explore attitudes, values, belief, satisfaction and social processes.[51]

Quantitative methodologies involve the use of statistical approaches and are based on quantifiable measurements of phenomena such as physical, behavioural, psychological, social and environmental factors. This is particularly useful for determining the impacts and outcome (effectiveness) of a program.[52]

Reflexivity is an ongoing process throughout a research project involving critical reflection on personal, interpersonal, methodological, and contextual factors that influence the study being conducted.[50]

Reliability in research is the consistency of a measure or method.

Replication in research is the process of repeating a study by independent researchers to verify the original result.

Representative sample is a small subset group that seeks to proportionately reflect specified characteristics of the target population.

Richness refers to whether data have context, personal meaning, emotional and social nuances, and layers of detail.[44]

Rigour in qualitative research refers to how researchers demonstrate the quality of their research. Research is considered rigorous or trustworthy when research can be confident in the study's methods, the data and its interpretation.

Sampling refers to the recruitment of a portion of the group from whom evaluators would like to collect data. In some cases, if the population is small, the whole population may be invited to the study.

Semi-structured interviews refer to interviews where some questions or topic areas are predetermined, while others are not. The interviewer uses the responses of the interviewee to provoke exploration of novel themes relevant to the research question.

Statistical power is the capacity of the statistical analysis to detect meaningful differences or change in outcomes given the natural variability of the underlying data.

Strengths-based stances promote a set of values that recognise capacities and capabilities rather than focusing on risk behaviours or problems.

Structured interviews refer to an interviews where the questions are predetermined in both topic and order.

Thematic analysis is a method for identifying themes and patterns of meaning across a dataset in relation to a research question.[36]

Triangulation is a way of combining data or findings from multiple sources to strengthen and validate findings for the same question.

Unstructured interviews refer to interviews where questions are not predetermined but follow on from the responses given by the participant.

Validity in research refers to the degree to which a study accurately measures or reflects what it claims to measure.

11. References

1. Liamputtong P. *Handbook of Research Methods in Health Social Sciences*. Singapore: Springer; 2019.
2. Leavy P. *The Oxford handbook of qualitative research*. USA: Oxford University Press; 2014.
3. Centre for Epidemiology and Evidence. *Planning and Managing Program Evaluations: A Guide*. Evidence and Evaluation Guidance Series, Population and Public Health Division. Sydney: NSW Ministry of Health; 2023.
4. Centre for Epidemiology and Evidence. *Preparing and appraising evaluation reports: A checklist*. Sydney: NSW Ministry of Health; 2020.
5. Centre for Epidemiology and Evidence. *Study Design for Evaluating Population Health and Health Service Interventions: A Guide*. Evidence and Evaluation Guidance Series, Population and Public Health Division. Sydney: NSW Ministry of Health; 2023.
6. NSW Treasury. *NSW Treasury Policy and Guidelines: Evaluation (TPG22-22)*. Sydney: NSW Treasury; 2023. Available from: www.nsw.gov.au/sites/default/files/noindex/2025-07/tpg22-22-evaluation-guidelines.pdf
7. NSW Treasury. *Evaluation Policy and Guidelines workbooks and resources*. Available from: www.nsw.gov.au/nsw-government/public-sector/financial-information-for-public-entities/centre-for-economic-evidence/nsw-government-investment-framework/evaluation-guidelines#toc-evaluation-resources
8. Peters S. Qualitative research methods in mental health. *BMJ Ment Health* 2010; 13(2): 35-40.
9. Florczak KL. Best Available Evidence or Truth for the Moment: Bias in Research. *Nurs Sci Q* 2022; 35(1): 20-4.
10. Maher C, Hadfield M, Hutchings M, de Eyto A. Ensuring Rigor in Qualitative Data Analysis: A Design Research Approach to Coding Combining NVivo With Traditional Material Methods. *Int J Qual Methods* 2018; 17(1).
11. Godwin A, Benedict B, Rohde J, Thielmeyer A, Perkins H, Major J, Clements H, Chen Z. New Epistemological Perspectives on Quantitative Methods: An Example Using Topological Data Analysis. *Stud Eng* 2021; 2(1): 16-34.
12. Leppink J. Revisiting the quantitative-qualitative-mixed methods labels: Research questions, developments, and the need for replication. *J Taibah Univ Med Sci* 2017; 12: 97-101.
13. Muijs D. *Doing Quantitative Research in Education with IBM SPSS Statistics*. Los Angeles: Sage Publications; 2022.
14. Hanson CS, Ju A, Tong A. *Appraisal of qualitative studies*. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2019.
15. Galdas P. Revisiting Bias in Qualitative Research: Reflections on Its Relationship With Funding and Impact. *Int J Qual Methods* 2017; 16(1).
16. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Med Teach* 2022; 7: 1-11.
17. NSW Treasury. *Evaluation Workbook III. Evaluation Plan: Design the evaluation*. Sydney: NSW Treasury; 2023.
18. Cranney L, Wen LM, Xu H, Tam N, Whelan A, Hua M, et al. Formative research to promote the Get Healthy Information and Coaching Service (GHS) in the Australian-Chinese community. *Aust J Prim Health* 2018; 24(3): 248-55.
19. Manby L, Dowrick A, Karia A, Maio L, Buck C, Singleton G, et al. Healthcare workers' perceptions and attitudes towards the UK's COVID-19 vaccination programme: a rapid qualitative appraisal. *BMJ Open* 2022; 12(2): e051775.
20. McGill B, O'Hara BJ, Phongsavan P, Bauman A, Lawler L, Grunseit AC. "I'm Still on Track": A Qualitative Exploration of Participant Experiences of a Weight Loss Maintenance Program. *Healthcare (Basel)* 2020; 8(1): 21.

21. Doyle L, Brady A-M, Byrne G. An overview of mixed methods research–revisited. *J Res Nurs* 2016; 21(8): 623-35.

22. National Health and Medical Research Council. National Statement on Ethical Conduct in Human Research 2007 (Updated 2018). Canberra: Commonwealth of Australia; 2018.

23. Aboriginal Health & Medical Research Council of NSW. *NSW Aboriginal Health Ethics Guidelines: Key Principles*. 2023.

24. NSW Ministry of Health. *Quick Guide on Undertaking Appropriate Aboriginal Health Research.* North Sydney: NSW Ministry of Health; 2020.

25. Crane M, Bauman A, Lloyd B, McGill B, Rissel C, Grunseit A. Applying pragmatic approaches to complex program evaluation: A case study of implementation of the New South Wales Get Healthy at Work program. *Health Prom J Aust* 2019; 30(3): 422-32.

26. Denzin NK LY, Smith LT. *Handbook of Critical and Indigenous Research Methodologies*. London: Sage Publications; 2008.

27. Rigney L-I. Internationalization of an Indigenous Anticolonial Cultural Critique of Research Methodologies: A Guide to Indigenist Research Methodology and Its Principles. *Wicazo Sa Rev* 1999; 14(2): 109–21.

28. Wilson S. *Research is Ceremony: Indigenous research methods*. Fernwood Publishing; 2020.

29. Lowitja Institute. Indigenous Data Governance and Sovereignty. 2023. Available from: www.lowitja.org.au/wp-content/uploads/2023/10/328550_data-governance-and-sovereignty.pdf

30. Bessarab D, Ng'Andu B. Yarning about yarning as a legitimate method in Indigenous research. *Int J Crit Indig Stud* 2010; 3(1): 37-50.

31. Atkinson P, Baird M, Adams K. Are you really using Yarning research? Mapping Social and Family Yarning to strengthen Yarning research quality. *AlterNative* 2021; 17(2): 191-201.

32. Kennedy M, Maddox R, Booth K, Maidment S, Chamberlain C, Bessarab D. Decolonising qualitative research with respectful, reciprocal, and responsible research practice: A narrative review of the application of Yarning method in qualitative Aboriginal and Torres Strait Islander health research. *Int J Equity Health* 2022; 21(1): 1-22.

33. Ungunmerr-Baumann M-R, Groom RA, Schuberg EL, Atkinson J, Atkinson C, Wallace R, et al. Dadirri: An Indigenous place-based research methodology. *AlterNative* 2022; 18(1): 94-103.

34. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm Policy Mental Health* 2015; 42: 533-44.

35. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018; 52: 1893-907.

36. Clarke V, Braun V. *Successful qualitative research: A practical guide for beginners*. London: Sage; 2013. p. 1-400.

37. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006; 3(2): 77-101.

38. Braun V, Clarke V, Hayfield N, Terry G. *Thematic Analysis*. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2019.

39. Grunseit AC, Gwizd M, Lyons C, Anderson C, O'Hara BJ. Polite, professional, practical: What drives caller 'satisfaction' with the New South Wales Quitline, Australia. *Drug Alcohol Rev* 2018; 37: S223-34.

40. Braun V, Clarke V. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *Int J Transgend Health* 2023; 24(1): 1-6.

41. Berry NJ, Gribble KD. Health and nutrition content claims on websites advertising infant formula available in Australia: A content analysis. *Matern Child Nutr* 2017; 13(4): e12383.

42. Miller CL, Brownbill AL, Dono J, Ettridge K. Presenting a strong and united front to tobacco industry interference: a content analysis of Australian newspaper coverage of tobacco plain packaging 2008–2014. *BMJ Open* 2018; 8(9): e023485.

43. Prior L. *Content analysis.* In: Leavy P, editor. The Oxford handbook of qualitative research. USA: Oxford University Press; 2014. p. 359-79.

44. LaDonna KA, Taylor T, Lingard L. Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Acad Med* 2018; 93(3): 347-9.

45. Fetters MD, Curry LA, Creswell JW. Achieving Integration in Mixed Methods Designs — Principles and Practices. *Health Serv Res* 2013; 48(6): 2134-56.

46. Steinmetz-Wood M, Pluye P, Ross NA. The planning and reporting of mixed methods studies on the built environment and health. *Prev Med* 2019; 126.

47. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007; 19(6): 349-57.

48. Korff JD. Deep listening (dadirri): Creative Spirits; 2023. Available from: https://www.creativespirits.info/aboriginalculture/education/deep-listening-dadirri

49. Keikelame M, Swartz L. Decolonising research methodologies: lessons from a qualitative research project, Cape Town, South Africa. *Glob Health Action* 2019; 12(1): 1561175.

50. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Med Teach* 2023; 45(3): 241-51.

51. Fry M, Curtis K, Considine J, Shaban RZ. Using observation to collect data in emergency research. *Australas Emerg Nurs J* 2017; 20(1): 25-30.

52. Bauman A, Nutbeam D. *Evaluation in a nutshell: a practical guide to the evaluation of health promotion programs* (3rd ed). Sydney: McGraw-Hill; 2023.