5.0

Evaluation methods

Delivered by
West Yorkshire Combined Authority | Tracy Brabin Mayor of West Yorkshire
West Yorkshire Violence Reduction Partnership
ECORYS Answering tomorrow's challenges today

5.1 Process evaluation

Generally, process evaluation seeks to capture the experiences of those delivering the intervention as well as those experiencing the intervention (i.e., young people, communities) to identify what is working well and any aspects that are, or may prove to be, problematic. Various methods are used to assess how an intervention is working and why it might work better in a different context. The task at this stage is to select the research methods most appropriate for answering your research questions. Process evaluations predominantly use qualitative primary methods,[29] including:

- Interviews
- Focus groups
- Observations
- Open-ended questionnaires
- Creative methods (i.e., photographs, videos, diary entries).

You will most likely also collect secondary data as part of a process evaluation, for example, conducting a document review or analysing project monitoring documents. It is important to uphold the ethical considerations laid out in Section 3 when conducting all of these research methods.

Sampling for qualitative methods

To begin with, it is important to consider who will be involved in your research. Depending on your research questions, you may wish to speak to staff at different levels within your organisation or project, intervention participants, members of the community, or funders. It is often not practical to involve everyone in your project within the research, so it is necessary to select a group of people (a 'sample'). Sampling methods which tend to be used in qualitative research include:

- **Purposive sampling**, whereby you choose respondents because they have a particular characteristic that you are interested in.
- **Snowball sampling**, whereby you ask one contact to put you in touch with others, and then they may also put you in touch with others (the sample size therefore 'snowballs' and becomes larger through word of mouth).
- **Convenience sampling**, which involves selecting people who are available.[30]

For qualitative data collection, the goal is to ensure that you capture the richness of views and opinions of the people involved; or to illustrate specific stories or contexts through in-depth case studies. There are no guidelines for how many respondents should be selected – it depends on the scale of your project and evaluation. It is important to understand that the objective with qualitative research is not to be able to produce findings that are statistically representative, but rather, to understand the range of views on a particular topic and to provide deeper insights which cannot be quantified.

29 It is possible to include quantitative methods such as surveys as part of a process evaluation.
30 https://golab.bsg.ox.ac.uk/documents/Good_Practice_Guide_-_Commissioning_Evaluations_Final_feb_14.pdf

Reaching **saturation point** (the point at which the addition of new data doesn't add anything new to the findings) is a good indication that the breadth of views and opinions has been captured.[31]

Interviews

Interviews are conversations between an interviewer (researcher) and an interviewee, (sometimes called a respondent or participant) in which questions are asked to gain information. Your research questions will guide who you will aim to speak to and the questions you will ask during the interview. The goal of interviews is to collect data and narrative information to better understand the respondents' unique perspectives, opinions, and world-views; the **lived experience** of other people is at the heart of this approach. Interviews invite the participant to make sense of their own experiences and to share these experiences with the researcher.[32] There are some important practical considerations for interviews here.

There are three types of interview:

- A **structured** interview is one with a set list of pre-planned questions. There is no scope for probes, and the conversation follows a specific pattern.
- A **semi-structured** interview, in which there are some guidelines around questions to ask (in the form of a list or 'topic guide'), but there is also flexibility around where the conversation can go. Researchers can ask follow-up questions and steer the conversation depending on the interviewee's answers to the pre-set questions.
- **Unstructured** interview, which can be described as 'conversations with purpose in mind' but without a set plan.[33] Whilst the researcher should keep the conversation on-track with the subject matter in mind, there are no set questions.

Interview questions can be **open-ended**, which encourage the interviewee to give their own answer, or **closed**, in which there are a limited set of possible answers (including 'yes' or 'no' answers). Open-ended questions are favourable in interviews, in order to gain more detail.

To begin with, use your research questions to think about what you would like to ask during the interview. It is not the case that research questions and interview questions are the same, however, there will be some overlap. The goal is to 'tap into' the respondents' experiences and perspectives, which requires asking questions that encourage the interviewee to talk and 'open up'. It is useful to formulate interview questions with the view that the interviewee should speak more than the interviewer during the conversation.

31 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/HMT_Magenta_Book.pdf

32 https://www.imperial.ac.uk/education-research/evaluation/tools-and-resources-for-evaluation/interviews/best-practice-for-interviews/

33 https://www.questionpro.com/blog/types-of-interviews/

These pointers may be useful when developing interview questions:

- **Word the questions so that respondents will answer completely and honestly.** Do not use 'leading' questions, for example, 'It is more difficult to engage young men than young women, isn't it? What do you think?'. Instead, ask 'Have you experienced any differences in engagement levels between different groups?'.
- **Develop probes that will give a more detailed response.** Simple probes might be 'would you like to say any more about that?' or 'could you please elaborate?', or they could relate directly to the question (see some examples in Figure 10 below).
- **Begin the interview with 'warm-up' questions.** This can be anything from 'Please tell me about your background and how you got involved with the program?' to 'Could you tell me about your interests and hobbies?'.
- **Think about the flow of the conversation** – what topics should come first, and what flows on 'naturally'. It is useful to consider embedding sensitive topics in the middle of the interview, rather than beginning or ending with them (although note that sensitive issues may arise at any stage, because the point of interviews is to encourage people to 'open up'. This is dependent on the topic area and respondent).
- **End the interview on a positive note**, leaving the respondent feeling empowered, listened to, and glad they took part.[34] You could ask if they have any further comments and let them know how helpful and insightful their comments have been. It is helpful to 'cool off' by asking if they have plans for the rest of the day, or even to talk about the weather.

Your interview questions will depend on what type of interview you are doing. Below is an example of the thought process (rationale) when developing interview questions with a research question in mind.

34 https://sociology.fas.harvard.edu/files/sociology/files/interview_strategies.pdf

This example is from a semi-structured interview with project delivery staff about VRU-funded interventions:

Figure 10: An example of the thought process behind developing interview questions with a specific research question

Research question	
Did the project reach the intended population and, if not, why?	
Interview question	**Rationale**
How would you describe the intended participant group for your intervention?	This will allow you to explore what project staff understood the target group to be, and some of the context around the choice of target group. It may also encourage them to express challenges or successes in engaging certain groups
How does the intervention reach and engage the intended target group? Probe: ▮ Is this effective? In what ways? ▮ To what extent does it reach those considered most at-risk?	Here you can begin to understand the ways in which the target population is reached and, from the project staffs' perspectives, the strengths and weaknesses of this approach. You could encourage them to elaborate on whether those most at-risk have been reached, and whether the approach was effective for all. This will help you to understand why outputs (and in turn, outcomes) have or have not been met for certain groups.
What has enabled/inhibited reaching and engaging with the intended target group?	This question will allow to you gain an understanding of barriers to reaching and engaging the target population. You could explore external factors, differences amongst groups, and how engagement could have been approached differently.

Focus groups

Focus groups are facilitated group discussions. Whilst interviews are often with one or two interviewees, focus groups use group dynamics to encourage people with shared experiences to share their views. Whilst there is no set size for the groups, having around 6-8 participants works well.

Focus groups are particularly useful if you want to want to use the group dynamic to develop new ideas which may not be possible through one-to-one interviews or other research methods, and if you want to consult with a range of people about a limited number of topics. They are, however, less appropriate if you want to explore confidential matters. There is a useful guide on setting up, conducting and analysing focus groups here.

Open-ended questionnaires

Questionnaires offer a systematic way to collect information from individuals or groups. Whilst surveys are the main way of collecting information in quantitative research, including qualitative, open-ended questions can allow you to gather data that is not easily quantified. This includes attitudes, beliefs, habits, and challenges. Open-ended questions can allow respondents to express the thought process behind survey responses and add context to quantitative data. Examples of open-ended questionnaire questions include:

- Please describe any challenges your project faced in relation to engaging young people.
- How would you describe the key activities offered to the young people involved?

Observation

Observations involve turning what you see and hear into evidence for research purposes. It allows you to assess a project or situation in real-time, seeing the activities and outputs laid out in a ToC in practice. This can provide a good understanding of the intervention being evaluated, from first-hand experience. Often, observational data is used in conjunction with other approaches as part of an evaluation design. For example, interviews, focus groups, or questionnaires might be used together with observations to build a more complete assessment of a certain activity.

The key to using observational data as evidence in an evaluation is to take a systematic and consistent approach as you collect, organise, and analyse what is observed. The steps to using observational study for evaluation (explained in detail here) include:

- **Setting the scope.** Establish the focus and approach- are you interested in a particular type of activity? Will you undertake a structured or unstructured approach?
- **Prepare the tools and the people.** This includes organising whether you will take notes or record, and organising the team and participants.
- **Conduct the observation.**
- Organise and analyse the data.[35]

You may wish to combine the above methods and data to create **case studies**, which focus on a particular aspect of your intervention, such as a person, place, or element of delivery. With this approach, you can gain insights into how certain factors such as policy change, local context, or life experiences have impacted individuals and communities.[36] A case study approach might include conducting 5 interviews with practitioners and 5 interviews with beneficiaries of one activity or adopting various methods with one beneficiary to create a case study of their individual story.

There is a range of other approaches you might use to collect qualitative data for your evaluation. Here is a useful resource with details on creative methods for qualitative research.

35 https://education.nsw.gov.au/teaching-and-learning/professional-learning/pl-resources/evaluation-resource-hub/collecting-data/observation/step-4-organise-and-analyse-the-data

36 https://www.evalcommunity.com/career-center/case-study-evaluation-approach/

Analysing qualitative data

The analysis stage of your process evaluation is the point that you begin to answer your research questions. It involves finding patterns and themes in the data that you have collected, which will help you report on it effectively and use it to make decisions. Qualitative analysis relies heavily on interpretation, so it is helpful to run the findings by multiple people in order to make sure that interpretation is not biased. If resource allows, you could test findings with a 'steering group' of participants or other stakeholders.

It is important to use a systematic approach to analysis, following these steps:

- **Prepare the data:** Firstly, prepare the raw data for analysis by transcribing audio recordings of interviews or by writing up interview notes. You may want to organise the data by research question, or stakeholder type. Here is a useful guide on transcribing interviews. You may want to analyse the data manually or using software (i.e., Nvivo or excel).
- **Get to know the data:** Read and re-read the data. Start to jot down thoughts and interesting points.
- **Identify common codes from an initial review:** A 'code' is a brief description of what is being said in the interview or focus group data. Think of it as a word or a short phrase that captures the meaning of specific quotes- a description, not an interpretation. It might be something like 'trust between young people and youth worker' or 'improved self-esteem'.
- **Categorise codes under themes:** A theme is broader than a code and captures something important about the data in relation to the research purpose. Read through the data and identify recurring themes, by applying the same code to sections of the text that represent the same meaning (i.e., if a young person said 'I feel more confident in myself', that would be coded to 'improved self-esteem'). You can use sticky notes and coloured highlighters to visually separate themes if you are doing your analysis manually rather than using software.
- **Review and refine your themes:** Ensure that there are identifiable differences between your themes and that they are useful and an accurate representation of the data. Also, consider whether anything is surprising and identify new themes you might have missed.
- **Report on the findings:** Reporting needs to go beyond just describing your data and should include your own analysis to make an argument about the story you present. Think of this as telling a story about your data, using quotes to back up points.

There are two methods of coding when conducting thematic analysis:

- **Deductive**, through which you come to the data with pre-determined themes that you expect to find based on knowledge or evaluation questions (this is more common in evaluation).
- **Inductive**, which allows the data to determine your themes as they emerge (this takes longer but is more exploratory).

These steps are summarised from this guide to thematic analysis.

5.2 Designing surveys for your evaluation

Define the objectives and research questions

If you have followed the steps set out in section 4.2 (on mapping research questions to data sources) then you should have identified which, if any, research questions you will answer through a survey. You will need to have a clear idea of what you are trying to answer through the survey to avoid making it too long, getting poor quality data or never using the results.

Determine the sample and administration method.

The first steps in gathering data with a survey is to decide who your target audience is – that is, your sample – and what method you're going to use to ask your survey questions. Your sample and the method of administering the survey will likely be interlinked; your intended audience may affect how you choose to distribute the survey. Both your sample and your method should be proportionate to your evaluation budget and the sort of project you have delivered.

Sampling

When determining who your survey respondents should be, firstly you will need to choose your **population** and then choose a smaller **sample** of this population to distribute the survey to.

Population: The entire group you want to draw conclusions from.

Sample: A specific group from your population that you will collect data from.

Some sampling methods have already been discussed in this toolkit (see section 5.1). However, you will most likely survey a larger group than you would for qualitative interviews, so there are more sampling methods that could be used. Ideally your sample should be:

- **Representative** – your sample should represent the core demographics of your population – that is the total number of people whose views you are interested in. For example, if you are interested in people's outcomes from engaging in your project, then your population would likely be the number of people who have engaged with your project in a certain period of time. If you're interested in young people's experiences of crime, then your population might be all those aged 16-24 in your local authority area).
- **Chosen at random** – this helps to ensure a representative sample and eliminate any bias in selection. For example, those who are most engaged in a programme are more likely to respond to a survey but may respond in a different way to individuals who were less engaged. If it is possible to select a random group from your population, then you should.

In some cases, your population may be the same as your sample. This would be the case if you are only interested in those who receive support from your project, and it is feasible for you to survey everyone who has taken part. Otherwise, you will need to consider how many people should be in your sample. The size of your sample will vary according to the size of your population and how accurate you would like your findings to be – that is, how much confidence you will have that your data reflects the views of your whole population. You can find more information about confidence intervals in section 5.3 (statistical significance), or here.

Once you have determined these two factors, you can use a simple online sample size calculator to work out how many responses you need.

Remember, it's unlikely that everyone you send your survey to will respond. This means you will need to send out more surveys than the number of responses you need. There are many variables that can affect your response rate, but a good survey response rate tends to fall between 5% and 30%, and an excellent response rate would be 50%.

Choosing an administration method

The way you choose to share your survey will depend on the way you interact with your chosen group and the resources you have available. For example, if you have regular face to face contact with your chosen group this may be the most effective for getting a high response rate. Some options are:

Type	Advantages	Disadvantages
Face to face	▮ Likely to get better responses to open ended, more qualitative questions. ▮ Inclusive for those with low literacy levels. ▮ Can support e.g., younger respondents, or those with learning difficulties, by helping them to interpret questions. ▮ Higher response rate if sampling the public.	▮ High cost. ▮ Respondents might be less likely to express their true feelings.
Telephone	▮ Likely to get better responses to open ended, more qualitative questions. ▮ Inclusive for those with low literacy levels.	▮ High cost. ▮ Might not have current contact details. ▮ Hard to reach people who won't answer unknown phone numbers.

Type	Advantages	Disadvantages
Text message/ SMS	▮ Quick to deploy, meaning you receive fast feedback. ▮ Wide reach to a large number of people. ▮ Cheaper than telephone surveys. ▮ High response rates due to easy participation.	▮ Limited survey length. ▮ Lack of detail in both questions and responses; limited numbers of characters to convey information. ▮ Might not have current contact details.
Online	▮ Low cost. ▮ Responses can be automatically collated ready for analysis.	▮ Excludes those who can't use technology well. ▮ Assumes a certain level of literacy in your sample. ▮ Can be exclusive of those with English as an additional language. ▮ Risk of low response rate.
Paper based	▮ Potentially more accessible, particularly if posted or given directly to your sample.	▮ Less environmentally friendly. ▮ Assumes a certain level of literacy in your sample. ▮ Can be exclusive of those with English as an additional language. ▮ More time consuming (and potentially expensive) to distribute. ▮ More time consuming to collate responses for analysis.

Resources for online surveys:

- Microsoft forms – Free. The software will automatically analyse your data for you, but you can't download the graphs produced.
- Google forms – Free. The software will automatically analyse your data for you, and you can download the graphs easily.
- Survey monkey – Mainly free, but some features are paid. Allows for more sophisticated question types.
- Smart Survey – Free for surveys of up to 15 questions, and 100 responses per month. Provides template surveys to personalise.

Ethical considerations

The data protection act covers how any organisation uses personal information. It gives individuals the right to know how their data is being used or have data erased amongst other things. It also stipulates that information collected is:

- used fairly, lawfully, and transparently.
- used for specified, explicit purposes.
- used in a way that is adequate, relevant, and limited to only what is necessary.
- accurate and, where necessary, kept up to date.
- kept for no longer than is necessary.
- handled in a way that ensures appropriate security, including protection against unlawful or unauthorised processing, access, loss, destruction, or damage.
- You should also consider whether using a survey is the appropriate data collection method. Sometimes when you are collecting sensitive data it may be more appropriate to conduct an in-person interview where you can provide support to respondents if they need it.

Design the survey questions

The next step is to design the survey, ensuring the questions you ask align with your research objectives. The questionnaire can include a mix of closed questions (multiple choice, rating scales) and open-ended questions. You might also want to include some questions on personal characteristics or demographics, so you can compare results for different groups.

Type of questions	
Question type	**Description**
Multiple choice	Lots of types of multiple-choice ques-tions including Likert scales, multiple statement multiple choices and more.
Likert scale	Type of multiple choice for rating how much an individual agrees or disa-grees with a statement. Either a 5- or 7-point scale; it's better to keep it simple.
Open ended	Free text boxes to get richer answers or to follow up a closed question.
Dichotomous	Questions with two responses, normal-ly Yes/No or Agree/Disagree.

Top tips when designing survey questions	
Tip	**Description**
Don't use leading questions.	**Bad:** 'Do you agree that this project is the most effective initiative in the community?' **Better:** 'How effective do you think the service has been?'
Don't use loaded questions.	**Bad:** 'Don't you think it's unbelievable that other projects are failing while our violence prevention project is making a real difference?' **Better:** 'Do you think this project is making a difference?'
Use plain English and use simple and short language where possible.	**Bad:** 'What are your perspectives on the efficaciousness of this project in mitigating societal harm from violence?' **Better:** 'How effective do you think this project has been in preventing the negative impact of violence on society?'
Use consistent phrases.	Just stick to one phrase or the other, don't mix and match or duplicate meanings of questions. Example: 'Do you think this project is successful?' and 'Do you feel the project has achieved its objectives?'
Keep the survey as short as possible.	You will be less likely to get responses if you produce a long survey, and many survey technologies will tell respondents roughly how long it takes to fill out. If you need a long survey, you may want to consider an incentive to complete it, such as prize draws or vouchers.
Research other surveys.	There may be other surveys done by your organisation, or others available online, which provide model questions you can utilise (for example, for demographics).

Informed participation

Regardless of your method of distribution, you should make sure that respondents understand why they are being sent the survey and what the purpose of the data collection is. You can do this verbally in telephone or face to face interviews, and with a cover page or letter in online or written surveys.

As a minimum you should include:

- The organisation behind the study, including a contact name and details for any queries.
- Why and how the respondent was selected.
- The aims of the study.
- What will happen with the information they provide.
- Data protection measures.
- Where relevant, any potential benefits or harms that could result from the study (and signposting to further support if you're surveying on a sensitive topic).
- That taking part assumes consent.

Pilot the survey

Once the sample and method of administration are chosen, and the survey is designed, it is important to conduct a pilot test of the survey with a small group of respondents. This help to identify any flaws, such as unclear questions or technical issues and can avoid problems with data quality later down the line.

Data collection

Once the survey is finalised, you can go ahead with data collection. You should monitor the data collection process while the survey is live to address any issues or concerns that arise – for example, if you're getting a poor response rate, why might that be? What can you do to improve it?

Data analysis

The next section of the toolkit, and especially page 58, has steps you can follow for analysing and interpreting the results of your survey.

You should also give extra consideration to consent if any of your respondents are under the age of 16. If this is the case, it is important to gain consent from parents or carers, especially if any of your questions might be considered sensitive or potentially upsetting. Where parental consent is needed, you should make sure parents are also given sufficient information to help them make an informed decision. This article[37] explores some of the issues around gaining consent from young people in online research and sets out some suggestions for best practice approaches

37 Harris, J and Porcellato, LA Opt-Out Parental Consent in Online Surveys: Ethical Considerations. http://researchonline.ljmu.ac.uk/id/eprint/8588/

5.3 Outcomes evaluation

Outcome evaluation is a tool for assessing the effectiveness and impact of your intervention. It is an objective process that involves collecting and analysing data to determine whether the intervention achieved its intended goals and objectives, and whether the outcomes are meaningful and beneficial.[38] It could be the first step in an impact evaluation (see Section 5.4.) and/or combined with a process evaluation (see section 5.1.) to create a mixed-method evaluation.

Specific research questions to be addressed by an outcome evaluation could be:

- To what extent did the project reach its outcomes?
- How did beneficiaries' outcomes change over time?
- Did the types/scale of outcomes differ by group characteristics?
- Did those that engaged more experience greater outcomes?
- Were there any associations between outcomes?

It should be noted here that an **outcomes evaluation cannot attribute a change in outcomes to your project**. It is for this reason that an outcomes evaluation is generally combined with process evaluation, so that qualitative findings can be used to support and interpret findings from the analysis of outcomes data.

Collating data on outcomes

The first step for conducting an outcomes evaluation is to begin with a review of the outcomes laid-out in your ToC. Selecting the appropriate data collection methods and outcomes measures is best done with your ToC at the forefront of your mind.

As covered earlier, your outcomes should be all be SMART (see section 2.3) and you should have established how you are going to measure them. Ideally, you will have been monitoring and collecting outcome data or have an identified secondary data source from the start of the intervention. Ideally, quantitative data should be collated into an electronic data set to make analysis easier – Microsoft Excel is useful enough for this, and most survey software will allow data to be exported to Excel.

Example: In Section 2.1, we shared an example ToC for a Hospital Navigators project. In that example, two primary outcomes were:

- 'Victim recognises vulnerability/seriousness of current situation'.
- 'Alternative non-violent life choices (recognised and available)'.

Each of these outcomes could be measured through a survey of participants when they join the intervention which is then repeated when they leave (called a pre- and post- intervention survey). Results from the survey can be collated into a data set as shown in Table 3.

38 https://www.evalcommunity.com/career-center/outcome-evaluation/

Table 3: Collating outcomes data

	Demographic information			Outcome data	
	Age	**Gender**	**Previous conviction**	**Outcome 1** Victim recognises their situation	**Outcome 2** Alternative non-violent life choices (recognised and available)
Individual 1	25	F	Yes	Yes	No
Individual 2	23	M	Yes	No	No
Individual 3	19	M	No	Yes	Yes

Analysing quantitative data

Dealing with a large dataset can be overwhelming, and it can be difficult to pin-point what the numbers are telling us. The following steps will enable you to demonstrate the effectiveness of your program in a manageable and tangible way.

Data cleaning

First, you must ensure that your data is ready to use, which means ensuring that it is accurate and free from any errors. Consider:

- Correcting spelling mistakes and inconsistent entries.
  - Imagine that you are using software such as Excel to analyse your data. Survey respondents may have inputted various answers to a question on their sex, i.e., 'Female', 'F' or 'Fmeale'. The software will not recognise that these entries mean the same thing, so you must go through the data and ensure that entries are consistent and free from typos. This can be a time-consuming task, but it cannot be avoided.
- Check for missing data and decide how to approach this, e.g. will a record be removed if enough data is missing? It is especially important to create a consistent approach if you are going to use summary statistics (see page 61).

Data matching

If your project has managed to secure data sharing agreements and you have access to data from a third party such as the police, you may need to match this to your own data. Matching data means combining two or more data sets based on a **matching variable**. In the case of individual level data, the matching variable will be an individual participant. To match individuals, you could allocate a **unique code** for each one, anonymising them whilst still making it easy to match records together. Combining data sets in this way can help you piece together a more complete story for your participants.

Example: You have two data sets: one contains demographic data for your sample and another the outcomes. Under GDPR, you may be required to remove any data that could identify a person which is not required for analysis, like names. However, allocating ID codes can also help with any issues arising from typing errors (see page 58). You may assume that Sam and Samuel are the same person, but with common names it is possible there are duplicates, and you may not have enough information to distinguish between them.

Data set 1			
Name	**Unique ID**	**Age**	**Sex**
Samuel Smith	S9867	25	M
Jane Jones	P0987	19	F

Data set 2			
Name	**Unique ID**	**Outcome 1**	**Outcome 2**
Samuel Smith	S9867	Yes	Yes
Jane Jones	P0987	Yes	No

Once your data set is cleaned, matched and complete, you can move on to analysing the data.

Summarising your data

To perform statistical analysis, you would ideally have a 'large' data set. However, there is no set definition of small or large data sets. An arbitrary minimum of 30 rows has been set by statisticians, however the general rule for analysis is the **bigger the data set the better**. The larger the number of people you have outcomes for, the more confident you can be that your analysis represents your target population.

There is also more you can do with a larger data set in terms of presenting analysis. Table 4 outlines the different statistical measures you can use to summarise your data.

Table 4: statistical measures for summarising data

Measure	Description	Type of data[39]	Example
Frequency	The 'count' or number of responses in each category.	Numeric, categoric, ordinal	25 individuals achieved outcome 1.
Percentage	The proportion of people who give a particular response.	Numeric, categoric, ordinal	50% of individuals in the programme achieved outcome 1.
Mean	The arithmetic average.	Numeric	On average individuals spent 14 days a term in exclusion.
Standard deviation	The average spread of scores around the mean.	Numeric	The standard deviation for number of days excluded was 10 days, indicating there was a lot of variation in experiences.
Median	The 'middle value' (another measure of average).	Numeric	Individuals spent a median of 10 days a term in exclusion.
Mode	The most common response to a question (another measure of average).	Numeric, categoric, ordinal	The outcome most commonly achieved was outcome 1.
Range	The difference between the highest and lowest response value.	Numeric	The number of days individuals in the programme were excluded in a term varied from 0 to 25.

Tools for data analysis

These calculations can be done easily in a range of computer programmes.

- Excel: The analyse tab and pivot tables provide quick and easy ways to analyse data. There are many free courses to help you use excel such as on LinkedIn, or you can find some cheat sheets here.
- Other data coding programmes are available, such as R (free) or Stata (paid).

39 You can read more about the four types of data here: https://www.upgrad.com/blog/types-of-data/

Issues to consider in your analysis

Small data

For small data sets/samples, take caution when presenting your data. Percentages and means can be misleading. For example, if you had
5 individuals and one individual thought the programme was 'great' you could say:

- **Option 1:** 20% of individuals we supported thought the service was 'Great'.
- **Option 2:** One individual we supported thought the service was 'Great'.

Option 1 sounds much more positive than option 2, however it is misleading as it doesn't tell the reader this was only out of a sample of 5.

There may be valid reasons why you have a small sample (for example, you might be looking at one particular aspect of your intervention). If this is the case you can still report outcomes, but the richness of your evaluation is more likely to come from qualitative data.

Summary statistics

It's important when you present summary data that you do so without misrepresenting your results. For example, of 100 individuals supported by an intervention, we have data that tells us that all 100 went on to secure a job. However, we only have prison records for 70 of them. Out of the 70, 40 had been in prison. Consider the following statistics we could pull from this:

- **Option 1:** 40% of individuals who got a job at the end of the programme had been in prison. This would be incorrect, as we do not know whether the 30 people for whom we do not have records were in prison or not.
- **Option 2:** 57% of the individuals who got a job at the end of the programme had been in prison. While this is factually correct, it doesn't offer the reader details of what sample was used. In other sections you may wish to refer to the full sample of 100.
- **Option 3:** 57% of individuals (of the 70 individuals we have records for) who got a job at the end of the programme had been in prison. This option is the most accurate, providing the reader with sufficient detail (and thus, transparency) of where the figure has come from. If you feel this level of detail is confusing, you could add it as a footnote or in a methodology note in the annex of a report or presentation.

Looking for data trends

By now you should have calculated some key statistics for your data. You may decide this is sufficient for your evaluation (if so, you can move on to page 64 – presenting your data). However, you could also take your analysis a step further and consider looking into trends in your data. This could be trends over time, or it could be trends in different subgroups in your sample. By digging further into the data, you will be better placed to answer questions linked to your ToC, particularly in relation to how your activities impacted on different groups, or whether different activities led to different outcomes.

Consider the following questions:

- How do trends vary between different subgroups?
- What are the key differences since you last analysed this data?
- How have trends changes over time?

Sub-group analysis

Subgroup analysis splits individuals into subgroups based on a variable (for example, age group or gender). Subgroup analysis can be used to compare differences within or between groups, and is important to pick up any effects on particular groups that might be hidden or less visible in the wider data set.

Example: Highlighting differences in experience through sub-group analysis.

Table 5 shows a sample of satisfaction ratings for five participants, including some demographic data.

Table 5: Satisfaction by gender and age

	Gender	Age	Satisfaction with service
Individual 1	F	20	Great
Individual 2	F	17	Ok
Individual 3	M	30	Ok
Individual 4	M	33	Poor
Individual 5	M	31	Poor

For this data, the median age for the whole group is 30, and the mean is 26.2.

Whilst those figures give an overview of the individuals receiving support, it would be much more illustrative to say that females who received support are generally younger than males, with an average age of 19 compared to 32.

Differences between subgroups may be more difficult to pick out when you have lots of data, but using software that creates quick summaries can quickly break your data down. In Excel, you could try using pivot tables to create cross tabulations of your results.

Example: In exactly 10 clicks we were able to create a pivot table (Table 6) using data from the previous example. From this table we can quickly see that only females thought the service was great and that only males though the service was poor. We could also say that generally older respondents thought the service was less good than younger respondents. By analysing data in this way, it raises questions about why there is a disparity in experience between sexes; answering that question could form vital learning for future delivery.

Table 6: Pivot table of satisfaction, by age and gender

Row Labels	Average of Age
Great	**20**
F	20
Ok	**23.5**
F	17
M	30
Poor	**32**
M	32
Grand Total	**26.2**

Choosing subgroups to explore

The Wales VPU outcomes framework resource includes suggestions of what sub groups would be useful to look at for different outcomes. For example, for the indicator (or outcomes associated with) 'violence related injuries', it is suggested that disaggregation could usefully be done by age; sex; ethnicity; injury type; day/time; relationship to perpetrator; violence type; deprivation quintile.

You might be interested to **compare your data over time** to build up a bigger picture of how the programme has evolved, or to understand the length of support needed to achieve outcomes. This could be done in a few different ways depending on what your outcomes are and how they are recorded.

Example: You are evaluating your programme after two years of delivery. There are three outcomes we are interested in looking at. If we didn't look at trends over time, we could say something like: '*By Q8 of the programme Outcome 1 was achieved by four individuals, Outcome 2 was achieved by 5 individuals and Outcome 3 was achieved by 3 individuals.*'

	Outcome 1	Date achieved	Outcome 2	Date achieved	Outcome 3	Date achieved
Individual 1	Y	Q3	Y	Q1	N	-
Individual 2	N	-	Y	Q2	Y	Q7
Individual 3	Y	Q3	N	Q1	Y	Q8
Individual 4	Y	Q4	Y	Q2	Y	Q8
Individual 5	Y	Q4	Y	Q1	Y	-

However, we are missing an opportunity to comment on time varying trends within the data. By looking at the data over time, you could say something with more insight like '*By Q8 of the programme Outcome 1 was achieved by four individuals, Outcome 2 was achieved by 5 individuals and Outcome 3 was achieved by 3 individuals. There was also variation in when the outcomes were achieved; all individuals had completed Outcome 2 by Q2 whereas no participants achieved Outcome 3 until Q7.*'

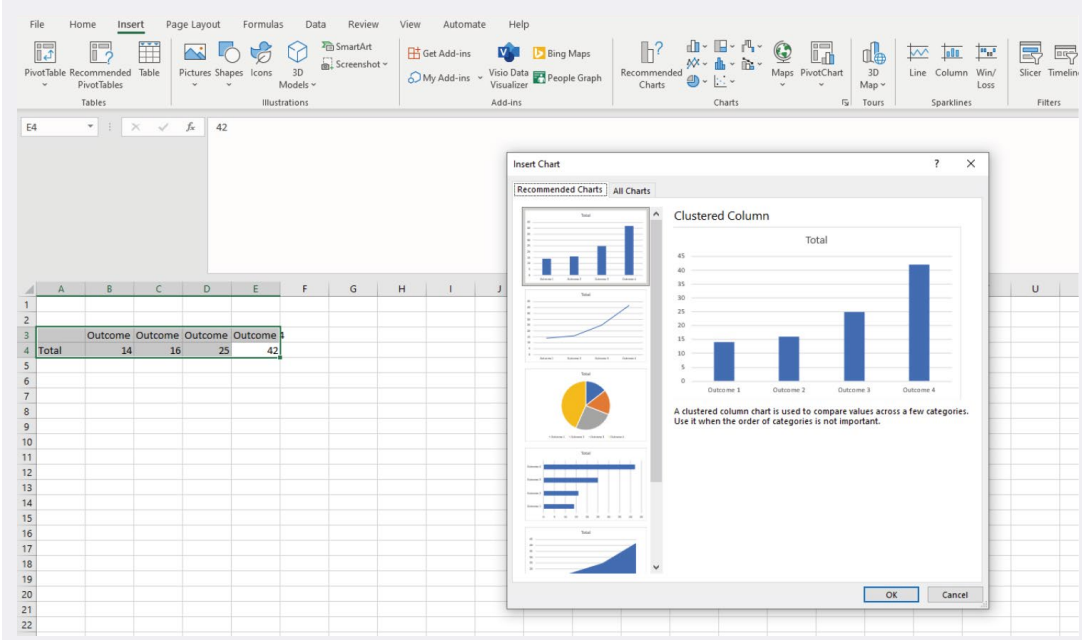
The next section will explore presenting your data graphically which is the easiest way to show any trends over time.

Presenting the data

From the processes outlined above, you should now have some outcome results, whether they are just summaries or whether you have identified more complex patterns in your data. The next step is deciding how to present your findings. Tables and charts are very useful to help present your results quickly and easily to another audience. If you are using basic data on excel, it will select suitable charts for you. It also gives tips on under what conditions to use the charts.

For example: if you wanted to show how many of each outcome your sample achieved, it would likely suggest a clustered bar chart. It also gives tips on under what conditions to use the charts. You can see an example of this in Figure 11.

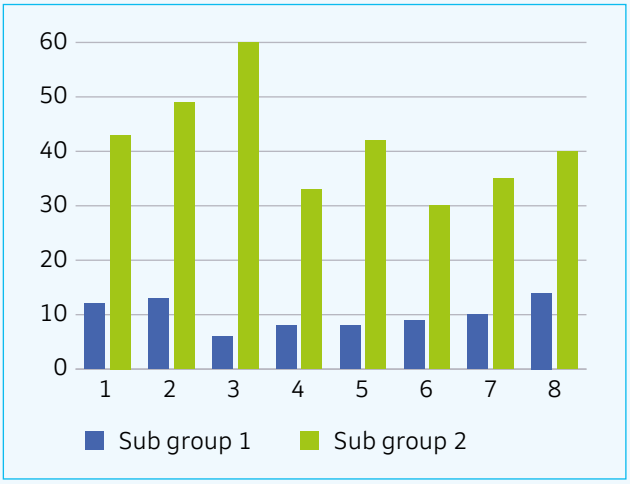
Figure 11: Choosing charts in Excel



You may want to present your analysis in graphs too, for example findings for trends over time or within subgroups. This may require some manipulation of your data to get it in the correct format to produce a graph. You can also make graphs using pivot tables.

Example 1: Outcomes achieved by different subgroups

Outcome	Sub group 1	Sub group 2
1	12	43
2	13	49
3	6	60
4	8	33
5	8	42
6	9	30
7	10	35
8	14	40



From the chart in Example 1, the reader can clearly see:

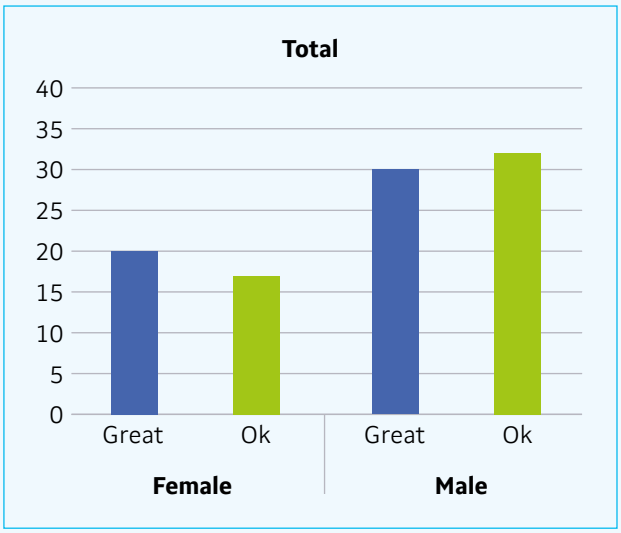
a) Subgroup 2 achieved more of every outcome than subgroup 1.

b) Outcome 8 was the most common outcome achieved by subgroup 1.

c) Outcome 3 was the most common outcome achieved by subgroup 2.

Example 2: Charts from pivot tables

Row Labels	Average of Age
Great	**20**
F	20
Ok	**23.5**
F	17
M	30
Poor	**32**
M	32
Grand Total	**26.2**



From the chart in Example 2, the reader can see the same conclusions we drew from this pivot table on levels of satisfaction by age and gender (see page 63, Table 6), but more easily and quickly.

There are several online tools and blogs to help you select graphs. One example is here.

After you have selected your charts and have your key figures, you should write up your analysis. This should include describing the results from all the steps described so far. Key figures should be summarised, and descriptions of graphs should be included picking out any interesting results. This can then be included in your evaluation report (see section 6).

Optional: Statistical significance

Statistical significance confirms that the result of your calculations did not happen 'by chance'. This can be done through statistical tests that provide us with a '**p value**' – a value that tells us the probability that the results we have seen have happened by chance or do in fact show a significant difference between groups.

T-tests

T-tests are used to compare the mean (average) of two or more populations or samples. It can confirm that two groups are statistically significantly different from each other. This could be used to test whether the cohort has significantly increased on an outcome. However, it's important to remember that this will not prove that your programme caused this change.

For example: A cohort of 1,000 young people took part in a sports programme where one outcome was to reduce the truancy rate amongst the group. This is measured by the number of unauthorised absences in a school term. The average number of absences across the group was 30 days before the intervention, and 25 after. A t-test compares these two values using the variation and sample sizes to confirm whether this change is significant or not.

There are several online calculators that you can copy your data into and it will perform the test for you. However, there are pre-conditions that the data must fulfil to use these tests effectively. Try this resource for more information and to access an online calculator.

Confidence intervals

Confidence intervals tell you how 'good' or accurate an estimation is, or how sure we are about the validity of our findings. This can be done as part of a T-test. For example, if there is a lot of variation in a data set, using a standard deviation test you can calculate a set of values which 95% of the expected observations should fall within. There is a more detailed explanation of confidence intervals here.

Should you wish to delve deeper into the statistical significance of your results, you can find more information in Graph Pad.

5.4 Combining findings from process and outcome evaluations

As explained in section 4, evaluation approaches should be thought of as complimentary. Process and Outcomes evaluations are rarely done in isolation and when paired together they can offer important insight into whether your project achieved change, and if so, how and why.

> *"An outcome evaluation tells you whether a program achieved its goals. A process evaluation tells you how and why."*[40]

An evaluation which combines different methods and approaches to collect data is called a **mixed-method evaluation**. For example, you could combine interviews and observations from a process evaluation with pre- and post-intervention surveys, and reviews of government statistics from an outcome evaluation. Whilst the two types of evaluation are geared towards different research questions, it is really useful to look at their findings together.

For example, imagine that your outcome evaluation research question is 'were alternative, non-violent life choices recognised by the young people involved? What impact has this had on their lives?'. Your process evaluation might explore which types of activity were particularly helpful for the young people, and why. You can combine the findings to gain an understanding of exactly how and why the intervention arrived at the observed outcomes, which is vital learning to take forward into future provision.

This approach can also be described as a **triangulation** of your findings across your different research methods, using a range of data sources and methods to ask the same question. You can begin to triangulate your data by thinking about trends in findings from different sets of data to see if findings converge or diverge. Where answers corroborate, you can be more confident in the validity of your findings.

Here is a useful resource on conducting mixed methods evaluations.

5.5 Methods for conducting impact evaluation

An impact evaluation, if done correctly, can attribute the impact on your desired target group and wider society to your project. In comparison, an outcome evaluation can only detail the change in outcomes in your target groups – it cannot attribute this change to your project.

Example: Outcome analysis: 25 individuals who took part in the programme did not commit an offence in the two years following the programme.

Impact evaluation: The programme *caused* 25 individuals to not commit an offence in the two years following the programme.

40 https://www.jbassoc.com/resource/whats-difference-understanding-process-outcome-evaluation/

5.0 Evaluation methods

Impact evaluation builds knowledge about what works, by assessing the change in outcomes which are attributable to the particular programme or policy. However, it is a difficult task attributing impact to your project. **Impact evaluation often involves using complex statistical methods and should only be used if you or your team have received training on the subject.** For most VRP projects where staff are conducting evaluation themselves, process and outcomes evaluation should be adequate to evidence if change has happened, and to understand why.

This toolkit will give a brief overview of impact methods, with links to other resources that will provide more detail on how to undertake one.

Why use impact evaluation?

- It is needed to establish whether programmes and policies have worked to improve outcomes.
- There is a growing trend to shift focus from inputs and outputs to outcomes, for better accountability and decision-making.

In his book *The Effect*,[41] Nick Huntington-Klein provides some examples of why it is important to be able to attribute a change to a policy or programme:

> *"We don't want to know if countries with higher minimum wages have less poverty, we want to know if raising the minimum wage reduces poverty. We don't want to know if people who take a popular common-cold-shortening medicine get better, we want to know if the medicine made them get better more quickly. We don't want to know if the central bank cutting interest rates was shortly followed by a recession, we want to know if the interest rate cut caused the recession."*

What research questions does impact evaluation answer?

There are many research questions that an impact evaluation can help to answer. Some examples of this are below:

- **What is the effect due only as a result of the intervention (and not other interventions)?**
- **To what extent can a specific (net) impact be attributed to the intervention?**
- **What would have happened without the intervention?**
- **How much of the impact can be attributed to the intervention?**

Ultimately impact evaluation will estimate the **causal effect** (impact) of:

- an intervention '**P**' (programme, treatment, policy) on
- an outcome '**Y**' (indicator, measure of success).

41 https://theeffectbook.net/

Unfortunately, this can be really challenging to do in social research. In an ideal scenario we would have a **counterfactual** to tell us what would have happened in the absence of P (programme, treatment, policy). This is normally in the form of a control group – that is, a group that did not receive P. The classic example in a medical trial would be those individuals who receive a placebo rather than treatment.

If we have a counterfactual and can 'control' for all other variables (referred to 'closing the back doors'; see more information on that here) then we could prove that the impact is due to P (programme, treatment, policy). However, it is difficult to form a counterfactual group for a social policy programme; it would be unethical to offer support to some people with identified vulnerabilities but not offer it to others. However, we will briefly explore different techniques to use to show impact by either inferring or creating a counterfactual.

Frequently asked questions about counterfactuals

Why don't we just select those who didn't take part in the project?

Selection bias: The reason a person did not engage might ultimately change the effect. For example, an optional course on how to get the best grades was offered to a classroom of school children. The children who took the course ended up with higher grades than those who didn't. However, it is likely that those children who took the course were more interested in getting higher grades and therefore might work harder, which would likely have an impact on their final grade. A correct approach here (though not necessarily ethical) would be to only allow half of the children who applied for the course to take it (at random) and then you could compare these groups.

Why don't we just compare before and after of the same group?

Contextual change: Other things may have happened in this time frame that we cannot control for and wouldn't be identifiable if you only study the treated group. For example, some of the children who took the course on how to get the best grades fell ill in the period before their exams. This impacted on their final grades and skewed the results for the cohort.

Data and project requirements

Not all programmes or policies are suitable for impact evaluation. Certain elements must be present to consider an impact evaluation:

- **Clear target outcomes/impacts** which should be SMART.
- **Intervention or project is consistently delivered** with no variation in approach.
- **Strong theoretical links between the project and desired impact:** there should be existing research and evidence that this type of programme would achieve the desired impact.
- **The *potential* for a counterfactual:** this can be hard to ascertain, but in some situations it might be particularly difficult (for example if a policy was rolled out across a whole nation).
- **Sufficient sample size:** as you need to use statistical tests for an impact evaluation, a minimum sample size is required. See here for more details on how to work out what this could be.

Impact evaluation methods

Generally, there are five key steps in the process of conducting an impact evaluation:

1. Feasibility study to establish whether the relevant data is available.

2. Research design, including sampling and establishing an analysis plan.

3. Randomisation (allocation to counterfactual groups) and data collection.

4. Analysis and testing assumptions.

5. Reporting.

The next few sub-chapters will explore some methods for impact evaluation design and analysis approaches.

Randomised Control Trials – The 'Gold standard' impact evaluation.

RCTs are **experiments** made of 3 main features: (1) control groups, (2) randomization, and (3) blinding.

1. **Control group:** a group not receiving treatment to be compared with the treatment group.
2. **Randomisation:** process of allocating participants randomly in treatment and control groups.
3. **Blinding:** procedures that prevent study participants, caregivers, or outcome assessors from knowing which participants were in treatment or control groups.

A typical example is a clinical trial, where one group is given regular treatment, while the other is given a placebo. The difference in outcomes is the net effect of the treatment.

RCTs are increasingly being applied in the field of social sciences despite the complexity of the approach. In the Youth Endowment Fund (YEF) toolkit, you can see examples of VRU projects that have been evaluated by RCTs.

As noted, not all programme or interventions are suitable for impact evaluation, and that includes RCTs. Table 7 explains when an RCT might be relevant and appropriate.

Table 7: Factors defining when to use an RCT

When is an RCT useful?	When is an RCT not useful?
▮ Interest in net impact. ▮ Intervention consistently delivered. ▮ RCTs planned from the start. ▮ Clear, measurable impacts.	▮ Interest in causal factors. ▮ High variation in how intervention is delivered. ▮ Quick answers are needed. ▮ Hard to measure or diverse results.
When is an RCT feasible?	**When is an RCT not feasible?**
▮ Possible to compare groups affected and not affected by the intervention. ▮ Possible to assign participants randomly. ▮ Large sample size.	▮ Not possible to compare groups affected and not affected by the intervention. ▮ Not possible to assign participants randomly. ▮ Small sample size.

Other Impact Evaluation methods (quasi experimental)

If an RCT is not possible or appropriate, there may be other impact evaluation methods you could apply in your evaluation. Below is an overview of other impact evaluation methods suited to different availability of data. You can follow the hyperlinks to resources providing more information on each.

Q: Do you have information on both treatment and control groups before and after the intervention?

- Difference-in-difference: Estimates impact by taking the difference between groups and before/after, controls for selection bias.

Q: Do you have individual-level data on known factors including participation for both groups?

- Propensity Score Matching (PSM): Matches the characteristics of treatment and comparator groups, allowing for better comparisons.

Q: Are individuals assigned to treatment and control group based on a cut-off point?

- Regression Discontinuity Design (RDD): Takes advantage of a programme 'threshold' (usually eligibility criteria to compare similar groups.

Q: Do you have data on trends before and after your intervention, but no concurrent control group?

- Interrupted Time-Series (ITS) analysis: when there is an expected 'interruption' due to a policy change.

Q: Can you use historical data to construct a 'clone' of a group receiving an intervention?

- Synthetic Control groups Method (SCM): Constructs artificial control groups based on a weighted average of existing comparator groups.

Theory-based impact evaluation

Theory-based impact evaluation can be used when using a counterfactual of any form is not possible. Instead, it tests whether the causal chains – the assumptions you made in your Theory of Change about the elements of your intervention that should lead to effects – are supported by strong evidence, and that other explanations for change can be ruled out. However, theory-based evaluation is less robust than using a counterfactual and is not able to conclude that change can be directly attributed to an intervention, but rather establish whether an intervention contributed to change.

Theory-based impact evaluation methods are recommended for use in very complex settings, for complex interventions, or when the intervention/policy/outcomes are designed to be adaptive (and so, aren't delivered in a consistent way with every participant).

Theory-based impact evaluation should broadly follow three main stages:

1

Develop a Theory of Change

Follow the process described in section 5 to develop your Theory of Change alongside stakeholders to accurately map any causal mechanisms. In a theory-based evaluation, alternative theories of how the change could have occurred should also be mapped.

2

Gather evidence

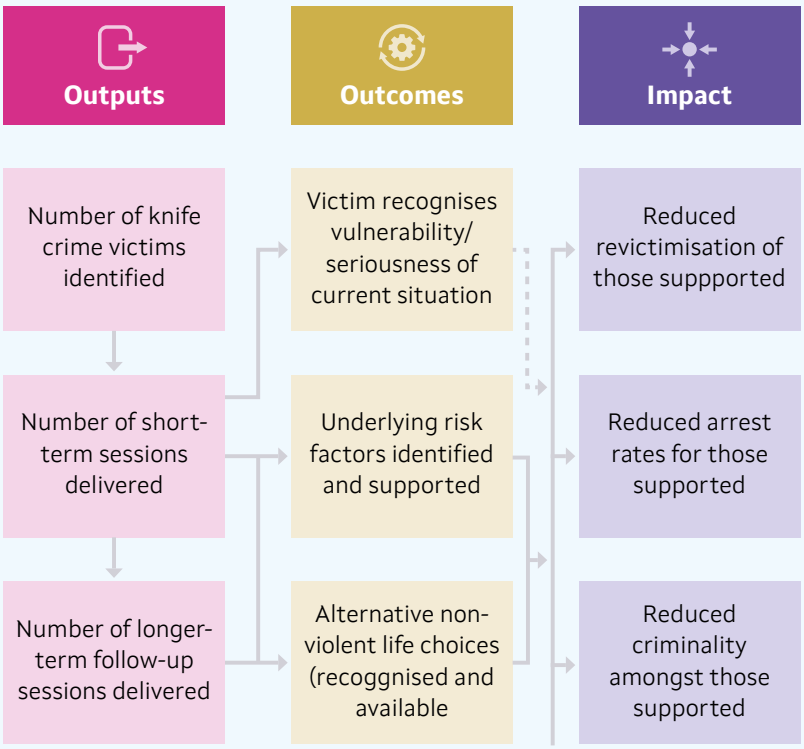
Evidence could be qualitative or quantitative and should be gathered with the purpose of testing causal pathways. The research questions should be developed accordingly.

3

Test causal pathways

Triangulate the evidence (see section 5.4.) and assess if, how, why, and when the intervention works.

Example: Looking back at the hospital navigators example ToC in section 2.1., we can see there is a causal chain between the output 'number of short term and long-term sessions delivered' and outcome 'non-violent life choices (recognised and available)'. To test whether this output led to the outcome, we could look at evaluation evidence which shows that sessions contributed to alternative non-violent life choices (such as exit interviews with participants, or follow up surveys), but also explore what other factors could explain changes in individuals' life choices. We could also test causal links from planned outcomes to impact and explore whether the take up of alternative life choices did lead to reduced criminality in the cohort.



There are different methods available to help weigh up whether your evidence is sufficient to conclude that your intervention contributed to the change. These are detailed in Table 3.2 of the Magenta Book.

Some other useful resources for theory-based approaches include:

- Contribution Analysis: an approach to understand the contribution of your program has made (or is currently making) to particular outcomes (as laid out in your ToC).
- Collaborative Outcomes Reporting.

5.6 Value-for-Money evaluation

You may have heard of value for money or 'cost benefit analysis' being used to explore the feasibility of capital or infrastructure projects. However, it can also be used in the evaluation of social projects or policies to express exactly what it says: how much value for money a programme or intervention provides. It does this by assessing the value of the outcomes or impact (not necessarily with a monetary value) and comparing this to the cost of the programme inputs. It may also consider what costs or benefits could have been incurred if another scenario took place (e.g. a different policy or programme). It can be used to answer research questions regarding whether the intervention was delivered efficiently or how much it cost to provide the outcomes delivered by the policy/project/programme.

Often, socially-focused projects aim to avoid or prevent a negative incident (such as a violent crime) taking place. Placing a monetary value on such outcomes can be complex, and generally involves assembling 'proxy' values of the cost to the public purse; in the case of violent crime, this may include costs of a hospital admission, an arrest, the criminal justice system and imprisonment.

Economic evaluation can be a powerful tool to communicate the value of your project. However, as with impact evaluation, **a value for money calculation should only be attempted if you or your team has received training**. The approach can be complex, and inadvertently creating misleading financial figures is likely to problematic for your organisation. This toolkit will briefly explain the different types of value for money evaluations, however it will not go into detail of how to undertake these methods. However, there are a number of resources that can help to assign a monetary value to outcomes if you do have access to an economic researcher:

- The Green Book: Government issued guidance on how to appraise projects.
- Manchester Combined Authority CBA tool: This includes a cost benefit analysis tool to input your project details into and also a generic data base that you can extract 'use values' from.

Requirements for assessing VfM

For a value for money assessment, you generally need:

- Outcomes or impact data for your cohort.
- Input cost data e.g. management information on staff costs, buildings costs, materials.

The following examples shows what kind of data we may need to conduct an economic evaluation of a programme.

Example: We have the following data for the individuals involved in the project.

Individual	Outcome 1: Not arrested violence for 12 months
1	Yes
2	Yes
3	Yes
4	No

Costs for the programme:

- **Staffing:** £15,000
- **Other overheads:** £5,000
- **Proxy costs for resources provided in kind (e.g. partner staffing, use of venues):** £5,000

We should also have a robust methodology in place for calculating the proxy cost to society of an individual avoiding an arrest for a violent crime.

Value-for-Money methods

There are different types of value for money assessments that can answer different questions:

- **Cost effectiveness analysis:** cost per outcome compares the cost of alternative ways of creating the same outputs. This is often used when cost benefit analysis cannot be undertaken as some valuations cannot be made but an outcome could still be counted (e.g. lives saved).

  Example: The programme in the example above cost £25,000 and achieved 3 crime-related outcomes at a cost of 25,000/3 = £8,333. This could be done in advance of delivery, with predicted outcomes, and compared to other programmes. Or it can be done after a period of delivery to assess the actual cost-effectiveness, and compare it to other programmes with the same outcomes.

- **Cost benefit analysis** assesses the impact of different interventions valued in monetary terms (where possible) by looking at all the costs and benefits of an intervention. Sometimes, the results are presented as a ratio of the value of benefits for every £1 spent on delivery. This type of approach is sometimes referred to as a Social Return on Investment analysis.

  Example: The cost to the 'public purse' of the crime 'violence without injury' is calculated as £6,480 and 'violence with injury' £15,353. The national VRU evaluation report published in 2022 applied these costs to outcome data generated by analysis of outcomes in a quasi-experimental impact evaluation,[42] and found the total estimated benefits of the programme were valued at £883 million. In this case, this represented a cost benefit ratio of 4.1 – that is, for every £1 invested in the programme, there was an estimated return of £4.10.

42 https://www.gov.uk/government/publications/violence-reduction-units-year-ending-march-2022-evaluation-report/violence-reduction-units-year-ending-march-2022-evaluation-report#section2