Scottish Government
Riaghaltas na h-Alba
gov.scot

Evaluation for policy makers A straightforward guide for policy makers

PUBLIC SERVICES AND GOVERNMENT

social research

Contents

Forward **4**

Chapter 1: What is evaluation? **5**
- What is evaluation? 5

Chapter 2: How does evaluation help to deliver policy? **6**
- The role of evaluation in policy making 6
- Evaluation as a standard part of policy work 6
- The National Performance Framework 7
- The importance of being objective 7

Chapter 3: How do I use evaluation effectively? **8**
- Finding an analyst 8
- Important principles for policy makers 8
- Issues to discuss with analysts 10

Chapter 4: What does the evaluation process look like? **12**
- At what stage should I consider an evaluation? 12
- Deciding what type of evaluation is required 12
- The step by step evaluation process 13

Chapter 5: What type of evaluation should I use? **15**
- Measuring impact 18
- Are control groups ethical? 19
- Natural Experiments 19
- Measuring outcomes and logic model evaluations 19
- What if I need to evaluation a new policy? 20
- Are evaluations all about outcomes? 20
- Are evaluations all about statistics? 20
- What is meant by quantitative and qualitative? 21
- What is monitoring data? 22

Chapter 6: What resources do I need to conduct an evaluation? **24**
- How much will an evaluation cost and how long will it take? 24
- Dealing with limited time and money 24

Chapter 7: What should I do with evaluation findings? 25
Communicating the results 25
Sharing the findings more widely 25
Using evaluation findings 25
Dealing with challenging findings 25

Evaluation Jargon Buster 27

Resources 29

Forward

I am delighted to introduce this user-friendly evaluation guide that was originally created specifically for policy-makers working in Scottish Government. However, it could also act as a useful reference for other agencies who aim to embed evaluation in their policy practices.

Evaluation has become firmly embedded in the work of the Scottish Government as it helps us to better deliver outcomes for the people of Scotland by understanding which policies work best and why, and how to learn and improve. It holds us accountable and ensures that policies benefit the people who they are designed to help. It also provides evidence for the public to better understand the decisions and activities of the Scottish Government, their impact and their value for money.

This guide covers a wide range of practical issues - from outlining the role of evaluation to examining different approaches to evaluation to helping you work out what to do with evaluation findings. It sets out key questions you might want to consider and to explore with analytical colleagues at every stage of the process. There are also a series of useful tips in boxes along the way.

I hope the guide succeeds in demystifying the topic of evaluation and convincing you of its benefits. As with many aspects of your work, the most useful evaluations result from collaborative working, in this case with policy makers, analysts and stakeholders working closely together from the early stages of policy development in a spirit of investigation and curiosity.

Dr Audrey McDougall
Chief Social Researcher

Chapter 1: What is evaluation?

A note about jargon

We've tried to keep jargon to a minimum but we've included a jargon-buster on page 27, in case you'd find it useful. If you see a word in **bold**, that means it's in the jargon-buster.

What is evaluation?

Evaluations use research methods to look at:

1. Whether policies have been implemented as intended (process evaluations) or
2. To measure whether outcomes have been achieved and how (impact evaluations) or
3. To assess if the benefits of the policy outweighed the costs (economic evaluations).

Evaluations should provide credible, impartial evidence about how your policy is working and enable the timely incorporation of its findings, lessons and recommendations into the decision-making process.

Importantly, evaluation is not about judging an individual or a team personally. Rather, it is an opportunity to:

- Check if a policy, service or intervention is being delivered as intended
- Measure the impact a policy, service or intervention is having
- For service users, stakeholders and practitioners to contribute their knowledge and views about how policies could be improved.

At the end of the process, evaluation provides feedback, recognises achievements that have been made, identifies ways of improving and supports evidence-based decision-making.

Chapter 2: How does evaluation help to deliver policy?

The role of evaluation in policy making

Evaluation helps to make sure policies are delivering the greatest benefit. Evaluations help us know if we are contributing to agreed outcomes, what's working, where to make improvements when they are needed, and how to allocate resources where they can make the most difference.

Evaluations help us understand which of our policies work best and why, and how to learn and improve. It also protects people by ensuring that policies benefit the people who they are designed to help, and don't have unintended consequences that cause harm.

Evaluation also holds us accountable and helps policy-makers and stakeholders to develop and implement the most effective policies to achieve the Government's objectives. If evaluation shows something isn't being implemented properly, or just isn't effective, it can help us make changes to improve the programme, or justify reallocating its resources to something more effective. Ultimately, evaluation is about using the results to improve policies, and by doing so, we improve people's lives for the better.

Evaluation as a standard part of policy work

Knowing how a policy is performing on the ground is a key component of the policy process. There are generally five main policy reasons for conducting an evaluation:

- There is an organisational or legislative requirement to do it.
- To learn lessons in order to improve the policy or identify best practice.
- To demonstrate accountability for spend.
- To assess whether there is enough evidence to roll out the policy more widely or provide further funding.
- To increase policy resilience (strong evidence will help conversations with critics).

The National Performance Framework

In Scottish Government, it is important to know how using evaluation can contribute to The National Performance Framework (NPF). The NPF, which was jointly launched by the First Minister and CoSLA in 2018, represents our collective vision of the kind of nation we want to be. 11 national outcomes describe this vision with a succinct purpose and set of values at its centre. The NPF is not merely a government framework but it belongs to the whole of Scotland where everyone has a part to play in delivering the national outcomes. The NPF also embeds 17 UN Sustainable Development Goals reinforcing our commitment to tackling the most challenging issues facing our planet.

There are 81 national indicators underpinning the outcomes that will help us track progress in achieving the outcomes over time.

The NPF functions on a 5 year cycle with the next review due to take place in 2023. To read more about the National Performance Framework, please go to (http://nationalperformance.gov.scot/)

The importance of being objective

Evaluation results are likely to suggest that a policy has strengths as well as limitations, so an evaluation should not be a simple declaration of policy success or failure. Evidence that a policy is not achieving all of its objectives can be difficult to accept, but it can also help you learn where to put limited resources.

Chapter 3: How do I use evaluation effectively?

Finding an analyst

In the Scottish Government, there are three analytical professions; researchers, statisticians and economists who can help policy to use evaluation. Although researchers tend to lead on evaluation, it is not the sole domain of researchers so we've referred to 'analysts' throughout the remainder of the guidance.

Important principles for policy makers

Although analysts are here to help, the effective use of evaluation relies on important principles that policy officials should adhere to – these are set out below.

Value the use of evidence in policy making

- Evidence-based policies are more likely to work. Use strong and consistent evidence from published evaluations to inform policy ideas and policy development.
- Start with the assumption that you don't know if a new policy will be effective until you have evidence.
- Conduct evaluations that help to improve policies, as well as evaluations that hold us accountable.
- Use the evaluation results to improve policy rather than keeping the report on a shelf.
- Appreciate the importance of investing in continuous data collection (monitoring) in order to underpin subsequent evaluations. In most cases it will not be possible to conduct an outcome or impact evaluation without monitoring or baseline data.
- Good evaluations are not only about numbers, especially if **sample sizes** are too small for numbers to make sense. Sometimes case studies can provide valuable insights into how policies are working and can be just as useful, or more useful, than costly and onerous data gathering exercises.

Work closely with your analysts and stakeholders from policy development stage

- Work with analytical colleagues (especially researchers) to make sure evaluation is considered at the earliest stage in policy development. The way the policy is structured and delivered and the way the data is collected by front line staff will make a huge difference to the quality of evaluation that can be done later, so it's always a good idea to check in with analysts from the very beginning of an idea.

- Have a clear understanding of what a policy is supposed to achieve, and share this with analysts.
- Decide on the questions you need to ask to determine if a policy is working, and consult with analysts about which evaluation methods will answer your questions.
- Support stakeholders to do self-evaluation, or ensure there is reliable data available for internal or contractor evaluations. Ideally provide adequate funds and/or support so stakeholders can conduct robust evaluations on our behalf.
- If you have to rely on stakeholders to recruit people into a pilot so it can be evaluated, then secure buy-in early on as the pilot recruitment process can take a considerable amount of time.

Balance spend with the need for robust evidence

- Good evaluations need resources to carry out. Think about whether evaluation costs can be built into budgets, and consider any financial implications for stakeholders if they are required to collect data at local level.
- Ensure that the scale and spend on evaluations is proportionate to the scale and spend on the policy. Analysts can help to give you an idea of the costs of different approaches, and to judge what is worth investing in based on the size and importance of the project.
- Build in more rigorous evaluations when the risk of policy failure has serious social and economic consequences.
- Using data already available (e.g. national surveys or administrative data), or supporting front-line staff to collect priority monitoring data on a continuous basis, can reduce the money you have to spend on full ‘snap shot’ evaluations.

TOP TIP

A good policy is evidence-based from the start

Evidence-based policies are more likely to work so use credible evidence to inform policy ideas and formulation. Ask analysts to look at the evidence-base or conduct a literature review. You can also use the various literature search functions to look for evidence on the topic of interest. The collection of robust published evaluations on your area of interest is sometimes called the ‘what works’ evidence base.

Issues to discuss with analysts

It is crucial to discuss evaluation with analysts during the early stages of policy development. Talking through the following points will help to get the planning process started:

1. Is there published, credible evidence that will inform policy ideas and the development of the policy?
2. Why is an evaluation required, who is it for, what do you want to know and when is it required?
3. What are your evaluation requirements at policy development stage (when policy objectives are being set)?
4. The rationale for the policy – what current issues or problems should it address? What impact is it supposed to have and on whom?
5. A **logic model**, so it is clear how policy objectives should be achieved, and how your policy should contribute to longer term outcomes and wider policy goals.
6. The key questions that you want answered about the policy.
7. The data that might be required to answer the questions and whether it is available.
8. The funds available for evaluation.
9. An evaluation approach (e.g. commissioning, internal, self-evaluation).
10. When you need the results (the timetable).
11. How will the messages from the evaluation be shared / disseminated and who with?

Don’t worry if you don’t know the answers to all of these questions right away, but giving them as much thought as possible in advance of discussing with analysts will make your initial discussion more productive.

TOP TIP

A good evaluation considers the interests of its audience

Be clear who the evaluation is for and what their particular interests are. This will determine the focus of the evaluation and the way you communicate the findings.

For example, practitioners might be more interested in evaluations that look at good practice and how findings can help them improve. They may also prefer a practical paper that focusses on practice recommendations rather than a long ‘academic’ report.

On the other hand, Ministers might be more interested in outcomes to see if the policy has achieved its aim. Also, think about how findings will be presented and shared with different audiences, so they have maximum impact.

Chapter 4: What does the evaluation process look like?

At what stage should I consider an evaluation?

The quick answer is at policy development stage because you need to make sure that early policy decisions don’t limit the evaluation you need to do later on. Even minor aspects of the way a policy is formulated can impact significantly on the ability to evaluate it. It is important to ensure that evaluation is considered and planned at the same time as the policy is being formulated so that the links between the policy design and evaluation options can be recognised and accounted for.

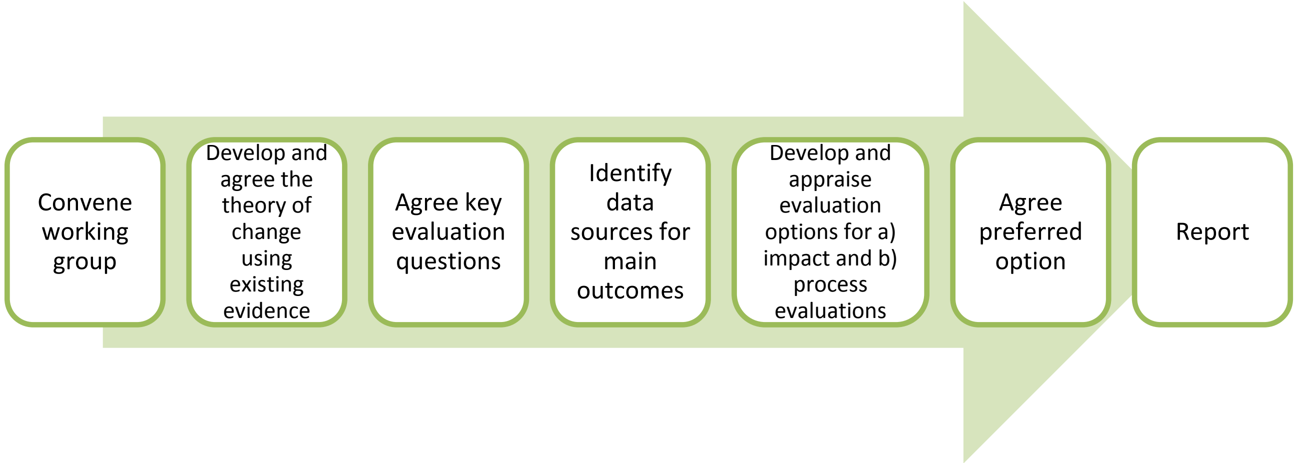
The importance of considering evaluation early: an example

Imagine you want to conduct a pilot to test a policy to decide if there is a case for rolling it out nationally so you decide to fund 3 pilots in comparatively low populated areas. Because the number of people using the pilot is relatively small, analysts tell you that it’s only possible to do a qualitative evaluation (using interviews). You’re disappointed as this won’t provide the information required to make decisions about national roll out. If you had discussed evaluation requirements with analysts at pilot development stage, then you could have considered expanding or extending the pilot to ensure the type of data you need to evaluate it could be collected.

Deciding what type of evaluation is required

For large, complex and expensive evaluations, it might be useful to conduct a formal ‘**evaluability’** assessment in collaboration with stakeholders, analysts and experts. The general process looks like this (Figure 1). Even if the evaluation is smaller scale, conducting a lighter touch evaluability assessment is always a very good idea. This process can ensure that evaluations are useful, and avoids committing significant resources to an evaluation if there is little realistic expectation of benefit.

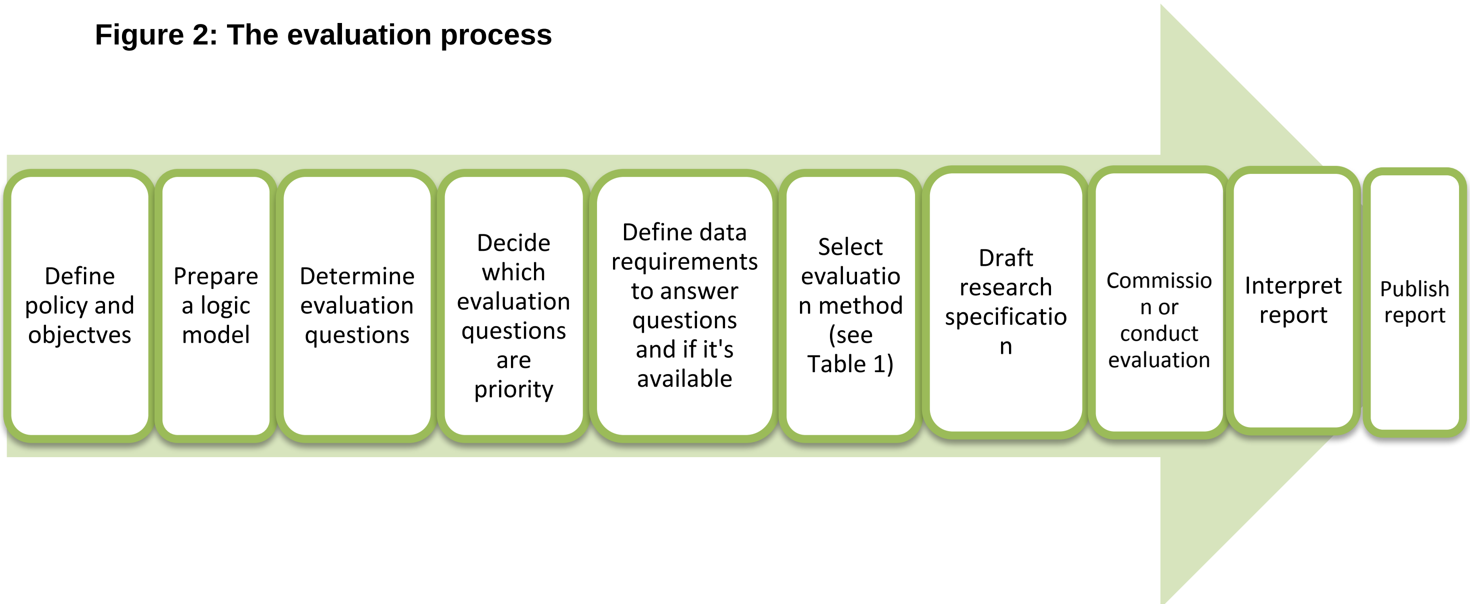
Figure 1: Evaluability assessment steps



The step by step evaluation process

If you all agree that conducting an evaluation is feasible and useful, then analysts, in collaboration with policy and stakeholders, are likely to follow a general step by step process that looks something like the process set out in Figure 2.

Figure 2: The evaluation process



TOP TIP

Combine evaluation with strategic planning

As you set longer term goals and a vision of the future, ask analysts to help identify ways to measure these goals and objectives and how you might collect, analyse, and use this information. This process will help ensure that your objectives are measurable and that you are collecting information that you will use. Strategic planning is a good time to create a list of questions you would like your evaluation to answer.

Chapter 5: What type of evaluation should I use?

The type of evaluation will be determined by what you and others want to know about your policy, the funds available, the time available and the feasibility of conducting the evaluation. Analysts can advise you and might suggest a range of possible evaluation methods. The table below shows one possible categorisation of the main types of evaluation that are used in the Scottish Government. They can be used in combination, and all contribute in their own right to better policy making. Notice how the type of questions you have will determine the type of evaluation you use. The strengths and limitations of each evaluation approach are also set out in the table.

Table 1 – How policy questions determine the type of evaluation used

YOUR QUESTIONS ABOUT THE POLICY	TYPE OF EVALUATION	STRENGTHS	LIMITATIONS
Process evaluations			
• Has the policy been delivered as intended? • Has the policy reached the intended target group? • What are the lessons learned and what could be improved?	**Process evaluation/ case studies**	Useful for identifying lessons learned and highlighting areas for improvement. Can be used to inform new policies where outcomes haven't had a chance to materialise.	Doesn't tell you about **outcomes** or the **impact** of the policy.
• How can we improve working practice or fix know problems on the ground e.g. in a health or education setting?	**Action or Participatory research**	Often lends itself to small scale studies. It can generate solutions to practical problems and can empower practitioners by getting them to engage with research and the subsequent activities.	Can be very time-consuming and biases can creep in. Difficult to generalise findings to other work contexts.

Table 1 – How policy questions determine the type of evaluation used (cc'd)

YOUR QUESTIONS ABOUT THE POLICY	TYPE OF EVALUATION	STRENGTHS	LIMITATIONS
Impact evaluations			
• To what extent have short term outcomes been achieved? • How many people have made progress from where they started (from a baseline)? • Has the policy achieved its own objectives?	**Outcome or 'Before and After' evaluation** (a data driven approach that uses a baseline measure as a basis for comparison but not a control group)	Useful if you want to know about the achievement of short term outcomes as it tells you how many participants have made progress on these.	You need to wait until outcomes have had a chance to materialise and there a few things to watch out for if you are considering this option. **Attributing** any pre-post changes to the policy is tricky which is why this design is recommended for shorter term outcomes only e.g. changes in knowledge before and after a test. Even if there is a difference in outcomes before and after the policy or intervention, without a **control group** we cannot be sure if the difference was *caused* by the policy or by other 'extraneous' factors. For example, the reoffending rate for participants after a domestic violence intervention may be lower than before simply because some of them were sent to prison after breaching the terms. The way in which participants are policed and willingness to report incidents may also be different before and after the intervention. You also need fairly large **sample sizes** if you want to test whether 'before and after' differences are **statistically significant** and also if you want to look at outcomes for subgroups of people e.g. older people or women. The interpretation of 'snap shot' outcome data can be challenging. For example, is 26% of people making progress on a particular outcome a positive finding or negative finding?
• Did the policy work as it was expected to work? How well has a policy been delivered? • To what extent have the short-medium term outcomes been achieved • What aspects of the policy	**Logic model evaluation** (this approach is based on testing your policy's 'theory of change')	Useful if you want to combine a process evaluation with a basic outcome evaluation. It can tell you how well the policy was delivered and also whether there was progress made on achieving outcomes. It can also measure the extent to which a single policy is making a *contribution* to longer	Same as an outcome evaluation.

delivered those outcomes? • Did certain types of people achieve better outcomes than other types of people?		term outcomes.	
• To what extent have the longer term outcomes been achieved? • Have those outcomes been achieved because of the intervention or because of other factors? • Was the policy initiative better than doing nothing or something else?	**Impact evaluation** (a randomised control trial or a matched comparison group) OR a **Natural Experiment** (where the experimental groups occur naturally such as men and women)	Useful of you want to know if the policy made a real difference.	Randomised Control Trials are regarded as the 'Gold standard' approach, but can be very difficult to do well in practice – especially if participant numbers are small, and/or a large enough control group is difficult to identify. The characteristics of the control group MUST be matched with the experimental group or any differences in outcomes could be due to unmatched samples. Unless an impact evaluation is combined with a process evaluation, it often doesn't tell you why or how the policy worked, just IF it worked.
Economic evaluation			
• Were the costs invested in the policy justified by the benefits? • Has there been an optimal use of resources to produce the outcomes?	**Economic evaluation or Value for money assessment**	Benefits and costs valued in monetary terms so they can be compared. Can tell you if the impacts are sufficiently valuable to justify further funding. Can answer questions such as: Have you minimised the cost of resources required to deliver services and achieve outcomes? Is the policy cost-effective?	Needs to be based on a robust impact evaluation. Limitations are the same as impact evaluation. It can take a while for benefits to be realised. Some benefits are difficult to value in monetary terms, e.g. social or environmental benefits.

Measuring impact

If your main policy question is ‘Does the policy work better than doing nothing or better than what was there before’, you should speak to analysts about doing an **impact** evaluation. This compares a **control group** (who did not experience the policy) with people who did experience the policy. Establishing causation (i.e. what causes what) is also a good reason to consider an impact evaluation - even if you discover that outcomes have changed over time, it doesn’t necessarily mean the policy has caused those changes. This is another reason some analysts suggest a control group - it can determine whether you can attribute outcomes to the policy or whether other factors may have had an influence.

A matched comparison or control group should be discussed very early in the policy development process to see if it is feasible because it requires sufficiently large sample sizes. A Randomised Control Trial is considered to be the ‘Gold Standard’ and is often used when the consequences of policy failure are serious.

TOP TIP

Be aware of self-selection bias

Randomisation can avoid self-selection bias that is introduced when participants choose whether or not to participate in the evaluation. This is also known as volunteer bias. Basically a group that chooses to participate is not equivalent to the group that opts out. This means the results of the evaluation can’t be generalised to the wider group of interest. Ask analysts how this can be avoided or dealt with.

Impact evaluations normally compare **quantitative** (numerical) data collected from the control group with the group that experienced the policy. However if the sample is too small for numbers to be meaningful, you could speak to analysts if it’s worth using a control group to compare people’s views and attitudes (**qualitative** data) with the group that experienced the policy. For example, if people who experience support from a policy initiative mention feeling satisfied more often compared with a control group who did not receive support, then it could indicate that your policy is achieving its objectives, even without numerical data. It is fairly unorthodox to use a control group in qualitative research, but it can be a partial solution to provide more meaningful information on how a policy is affecting its target group if numbers are small.

Are control groups ethical?

Some people may strongly object to the idea that some participants in the evaluation will not experience the policy or intervention just so that outcomes can be compared with those who do. This is often the case if the policy or intervention seems beneficial on the face of it. However, it is worth bearing in mind that until the policy is tested we have no way of knowing if it is beneficial or not. It could make no difference at all (a waste of valuable resources) or even be harmful. Some would go further to argue that it is more unethical to roll out an untested policy, especially if the consequences of policy failure were serious.

However, do not use a Randomised Control Trial or a matched comparison group if your sample sizes are too small. Your available sample size is often determined by the design and scale of the policy. An analyst can advise on whether a sample is large enough for Randomised Control Trials. If numbers in each group are too small the difference between your 'policy' group and your control group would have to be unrealistically large for the difference to be **statistically significant**. If you are unable to do an impact evaluation, this would mean that you cannot answer certain questions about the policy, for example, are the policy outcomes better than doing nothing.

Natural Experiments

A natural experiment is sometimes used if randomisation is considered unethical or difficult to implement. It is where individuals are 'naturally' assigned to the experimental and control conditions but the process governing the exposures arguably resembles random assignment e.g. families with 2 male children or 2 female children. Thus, natural experiments are observational studies and are not controlled in the traditional sense of a randomised experiment.

Measuring outcomes and logic model evaluations

If an impact evaluation is not feasible, then an **outcome** or **logic model** evaluation can be very useful alternatives, but you still need sufficient sample sizes to quantify outcomes (see Table 1). The beauty of logic model evaluations is that they combine a process evaluation (that will explore if a policy is working as intended) with an outcome evaluation (measures the extent to which outcomes have been achieved).

TOP TIP

Logic model evaluations can test if your policy is achieving its own objectives

Every policy has a 'theory of change' that underpins it. In other words, that doing X will achieve Y. If it is impossible to obtain a control group or you have small sample sizes, discuss whether a **logic model** evaluation would be useful. This type of evaluation will assess if outcomes emerged as was predicted when the policy was formulated. It will also assess the extent to which your policy is contributing to the achievement of longer term goals by measuring the achievement of associated shorter term objectives.

What if I need to evaluation a new policy?

In the early stages of a policy, a process evaluation (see Table 1) should tell you how well it's being implemented and delivered, which will highlight where improvements may have to be made before the policy is rolled out or given more funding. Only when outcomes or benefits have had a chance to materialise should you consider an outcome, logic model or an impact evaluation - but this should build on data collected in the early stages about implementation and delivery.

Are evaluations all about outcomes?

That's a good question. There has been a lot of focus on outcomes and impact lately, but the fact is, benefits or outcomes are unlikely to materialise unless your policy is evidence-based and is implemented and delivered to a high standard. This is why it is worth conducting a process evaluation or a logic model evaluation to make sure your policy is running as intended. Another issue is efficiency. For example, you're happy to discover that you policy is achieving some outcomes, but then a value for money evaluation (see Table 1) finds that the costs of the policy exceed the benefits achieved. On balance, you might conclude that it might be better to spend those resources on something else.

Are evaluations all about statistics?

No, not only about statistics. Once you've settled on which general method is best suited to your purposes, you can refine your specific questions and decide on the balance between **quantitative** and **qualitative** data you need to answer them.

What is meant by quantitative and qualitative?

Data can be quantitative, meaning it is measured as a number, or qualitative, meaning it expresses ideas or opinions in words. For example, suppose you want to know whether an intervention helped people to feel more confident. Here are some examples of approaches you might use:

Quantitative:

- Use a **validated scale or inventory** to score their confidence levels before the intervention and after. Compare the scores over time, and ideally with a con**trol group** as well.
- Ask participants a question like "Do you feel that the programme has helped to increase your personal confidence? YES / NO/ UNSURE" and compare the number of people who select each answer.
- This is particularly helpful if you want to be able to quantify how much people's confidence levels changed.

Qualitative:

- Ask a survey question like "how do you feel about your self-confidence since participating in the programme?" then analyse the answers people give to identify common themes.
- Use semi-structured interviews with programme participants to ask questions about their confidence and explore their answers in more depth follow up questions.
- This is particularly helpful if you want to understand how and why people's confidence levels did or did not change, and what they did and didn't find useful about the intervention.

TOP TIP

Statistics aren't the be all and end all

It is easy to be seduced by the apparent "certainty" and persuasiveness of statistics, but remember – they are only as good as the **methods** used to collect them and isolated statistics are sometimes prone to misinterpretation. Talk to your analysts about how data will be collected to make sure possible bias is minimised, and how they should be presented, explained and contextualised.

As explained above, if you want to be confident that a policy or intervention has had an impact on outcomes you need to collect quantitative data, use a control group who didn't experience the policy, compare their outcomes with the group that did experience the policy and find a statistically significant difference.

However, good evaluations are not only about numbers. Often the richness that comes from good quality case studies gives us a much deeper understanding of a situation's context and complexities, which might be even more useful to policy formation and improvement. They can tell us how a policy works or doesn't work.

To balance out their strengths and weaknesses, most evaluations will use a combination of quantitative and qualitative methods. This lets us get a robust descriptive picture through quantitative statistics, and to explore the "how and why" questions with more nuance through richer qualitative data.

What is monitoring data?

Without access to good data and evidence we cannot make informed policy decisions, or invest limited resources where they will have the biggest impact. If we have access to an existing database of relevant information it can save precious time and money as you may not need to commission a full evaluation. Also, a lack of reliable monitoring data can severely hamper standalone evaluations.

There are sometimes opportunities to change or add questions to national surveys. So it is worth considering whether existing data capture tools could ask the type of questions that would shed light on how your policy is performing.

TOP TIP

Make the best use of the data we have

'Snapshot' evaluations assess a policy at a given point in time. But analysing relevant information from existing datasets, for example from national survey data or administrative data can be highly cost-effective and you can revisit the data to see if attitudes and behaviours are changing over time.

It is worth being aware of some challenges with existing data. If we can resolve these issues with analysts and data custodians, then the evidence on how policies are working and for whom would be greatly improved.

Some common issues with existing datasets:

- Some data custodians experience technical issues in making their data easy to access – for example, if they are still running paper-based systems or older-style databases that have limited functionality.
- Data is often collected in local silos so it can be tricky to get a national picture or compare one area with another.
- Some data is hard to access especially if we don't own it. Data sharing agreements may need to be drawn up.
- Some of the data collected can't answer important policy questions. For example without unique identifiers and joined up systems, it's hard to track people through the education system, the health system or the justice system.

Chapter 6: What resources do I need to conduct an evaluation?

How much will an evaluation cost and how long will it take?

Well, it varies depending on the type of evaluation you need to do. The following two fictitious examples will give you some idea of each end of the typical range but most evaluations seem to fall somewhere in the middle of the two examples.

For example, say policy team X wanted to know if a project they had rolled out was working as intended or if it could be improved, so they commissioned a research organisation to analyse the basic user data collected by front-line staff, speak to project practitioners and some service users. This took two months to procure, the study ran for three months and could cost up to £20,000.

At the other end of the scale, policy team Y wanted to know if a national policy was working. This involved interviews with members of the public and people who were affected by the policy. The contractors also conducted a comparative experimental study using a control group. Because this method involved negotiating access to the public, and people affected by the policy - some of who may have been in difficult situations – these elements were complicated to set up and required specialist skills so the study took 18 months at a cost of over £100,000.

Dealing with limited time and money

There are other ways to evaluate policy performance if it makes sense for your particular policy. As we already mentioned, existing datasets such as administrative data or survey data could be the answer. If there is no existing data available, and your policy has national coverage, you could also pay for questions to be included in Omnibus or Panel surveys. This can give you a national ‘snapshot’ of public views or behaviours at different points in time (e.g. before the policy and sometime after roll-out) which might be more practical if you need a quick answer on how a national policy is performing.

TOP TIP

Check legislation or funding requirements early on

Sometimes legislation or funding agreements might specify who has to do the evaluation, for example if it is required to be an external body in order to guarantee impartiality. Make sure you check this before you start discussing options with your analyst.

Chapter 7: What should I do with evaluation findings?

Communicating the results

Although this is one of the last evaluation tasks, discuss upfront how the results will be shared. Most importantly, identify who your primary users are. Of course the primary user may be you and your policy team but your findings can be communicated to others for different reasons. For example, lessons learned from the evaluation can be helpful to stakeholders in the same field; or it may be worthwhile remoulding some of the findings into articles or stories to attract wider attention.

Sharing the findings more widely

Although a final evaluation report is important it is not the only way to distribute and communicate findings. Depending on your audience and budget, it may be important to consider different ways of delivering evaluation findings:

- Presenting findings at seminars and conferences
- Community events
- Tweets and other social media
- Sharing written briefings for different audiences
- Developing a short video version of findings (e.g. on Vimeo)
- Sharing stories and pictures from the evaluation (depending on what options you have used to gather data)
- Creating large posters or infographics of findings for display

Using evaluation findings

How the evaluation results will be used should be considered at the start as this will shape the evaluation brief. For example, if the aim is to use the results to adjust the policy then taking an improvement approach will make sure that the evaluation is focused on lessons learned. The report should also include suggestions for improvement as well as unearthing good practice. You should also consider how to respond to both positive and negative findings early on so you can prepare before the evaluation is completed.

Dealing with challenging findings

This issue highlights the value of taking a learning approach to all evaluations – nothing is a 'failure' if we agree to learn from evaluations and take action to improve. This also highlights the importance of rigour, for example obtaining decent sample sizes for pilots and using control groups. Better quality evaluations provide more justification for action.

Negative results are equally valuable as a way of identifying ‘what not to do’. Of course, the reasons for failure should be fully explored – was it a poor policy or poorly implemented?

- Talk through your evaluation requirements with analysts at policy development stage
- Make sure early policy decisions don’t unintentionally hamper the scope of an evaluation further down the line
- The evaluation method and the type of data collected will depend on what you want to know and what is feasible to do
- Evaluations can be as much about learning and improvement as they are about accountability
- Manage expectations from the start – evaluations can’t answer every question!

Evaluation Jargon Buster

Evaluation	The rigorous collection and analysis of information that determine the merit or worth of the policy.
Evaluability	A formal set of steps that assesses whether a policy can be evaluated in a reliable and credible fashion.
Outcome	The short-medium term changes or results that a policy or intervention should or has brought about. For example a change in attitudes or behaviour.
Impact	Impact is when you can determine if outcomes have occurred as a direct result of the policy or intervention. In a logic model, 'impact' can mean the long-term, cumulative effect of policies over time on what change they ultimately aim to contribute to such as change in the national crime rate or morbidity and mortality.
Logic model	A logic model is a graphic that explains how policy activities should lead to a series of outcomes that, in turn should contribute to an ultimate outcome or impact . It can be developed for any level of intervention – an event, a project, a programme, a policy, a strategy or an organisation.
Baseline	This is evidence you gather ideally before an intervention has been delivered. Therefore, if you gather similar information afterwards, you can compare the findings before and after the intervention and hence assess to what extent and whether the outcomes may have been achieved Note: Just because there is a change from start to finish doesn't necessarily mean that the policy has 'caused' that change.
Monitoring data	Data and information that is collected continuously during the lifetime of a policy or service. Data can be collected on user numbers, completion rates, activities and progress towards achieving outcomes. Monitoring data should be more relevant to practitioners and managers than evaluations as it can be analysed at any given point in time.
Validated	A validated questionnaire, scale or inventory has been tested to make sure it measures the constructs that it is supposed to measure e.g. 'anxiety' or 'motivation'. It has also been tested for reliability (the ability of the questionnaire to produce consistent results).
Sample size	Is the number of units from which data will be gathered in an evaluation e.g. the number of participants in a service who answer a questionnaire.
Control group	The control group is the 'untreated' group who don't experience the policy or intervention, but from which data is still collected. The 'treated' group do experience the policy or intervention. Having a control group means you can compare the results of the treated group with untreated group to see if there is a real difference between them. Another way of seeing the control group is as a baseline to compare groups and assess the effect of the policy or intervention.

Effect size	'Effect size' is a statistical term. It is simply a way of quantifying the size of the difference between two groups. For example, it can calculate the size of the difference between test scores in a group of people who experienced your policy and the test scores of a control group. If you have a small sample, the effect size (the size of the difference between the two groups) would have to be very large to have not occurred by chance (be statistically significant).
Statistical significance	Another statistical term. If a result (for example the difference in average test scores between two groups) is **statistically significant** then the difference is statistically large enough *not* to have occurred by chance or by fluke. We use a test to determine this.
Attribution	The process by which the *causes* of outcomes are explained.
Method	The particular way that an evaluation is carried out.
Scale or inventory	A list of items that are known to measure a particular construct e.g. level of confidence or anxiety.
Quantitative	Data is captured by counting and measuring things and analysed by making numerical comparisons. Quantitative approaches can allow us to generalise results from a sample to the wider population of interest.
Qualitative	Data is captured through a person's views and observations and reported in the language of the people interviewed.

Resources

More on evaluation methods
https://www.gov.uk/government/publications/the-magenta-book
http://www.betterevaluation.org/

How to do logic model evaluations (with examples of logic models)
http://www.gov.scot/Publications/2016/05/1967

Where third sector colleagues can access evaluation support
http://www.evaluationsupportscotland.org.uk/

Value for money in commissioning strategies: economy, efficiency, effectiveness and equity.
https://www.nao.org.uk/successful-commissioning/general-principles/value-for-money/assessing-value-for-money/

Economic evaluations
https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

© Crown copyright 2018
You may re-use this information (excluding logos and images) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/ or e-mail: psi@nationalarchives.gsi.gov.uk.
Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

The views expressed in this report are those of the researcher and do not necessarily represent those of the Scottish Government or Scottish Ministers.

This document is also available from our website at www.gov.scot.
ISBN: 978-1-78781-402-8

The Scottish Government
St Andrew's House
Edinburgh
EH1 3DG

Produced for
the Scottish Government
by APS Group Scotland
PPDAS502146 (12/18)
Published by
the Scottish Government,
December 2018

Social Research series
ISSN 2045-6964
ISBN 978-1-78781-402-8

Web and Print Publication
www.gov.scot/socialresearch

PPDAS502146 (12/18)