Social Policy Evaluation and Research Unit

superu

# Making sense of evaluation:

## A handbook for the social sector

**using Evidence for Impact**

# Contents

# Contents

# Introduction

## Who is this for?

This handbook is primarily designed for people funding or purchasing social services, and providers of social services in New Zealand who need to understand more about how to measure and understand the effects of their programmes or initiatives. This can help people who are working with:

- programme design (and funding applications)

- management of programmes through good monitoring and feedback, including continuous improvement

- reporting on outcomes

- understanding what works, when and why.

## What we cover in this handbook

The handbook has been developed to provide people new to evaluation with an overview of evaluation concepts and processes so that they are sufficiently equipped to work in this space (often, but not always, with a professional evaluator). In Module 4 we talk about how you can find an evaluator.

It is not designed to be the definitive guide – there are many sources of help and guidance already in existence. We refer to some of these throughout the handbook.

This handbook specifically relates to the New Zealand social sector and, as far as possible, avoids jargon and conceptual debates. However, it may be of interest to others given that the principles that are outlined are generally applicable to programmes and initiatives beyond the social sector.

**I'm a provider.** What do I need to provide to ensure that I can **demonstrate the effectiveness** of my programme?

**I'm a project manager.** I have to **commission an evaluation** and I want to know how I can talk to evaluators.

**I'm a programme manager.** I want to know **whether the programme I funded is effective,** and how robust the evidence is to support that.

**I'm a funder.** I want to be able to make decisions on how to **allocate funding** between programmes/providers and to get best value for my investment.

**I'm a policy maker.** I want to **know if my initiative is on track,** reaching the people it should, and having the outcome I want.

# How to use it

We have broken this handbook down into a number of modules, which look at some of the most important aspects of evaluation. Each one identifies what you should be able to get out of it, and links you to a range of complementary resources that are more detailed, provide more sector-specific examples, or are tools or frameworks that you might find useful.

---

## Why do we need it now?

A variety of reasons make it necessary for people to be more evaluation focused:

- being over-worked and under-resourced are a key incentive to better prioritise and apportion resources to where they are needed/make a difference

- scarce funding sources (e.g. government and philanthropic), which push you to better demonstrate the positive changes you're making in people's lives to access this limited funding

- personal satisfaction in knowing that you're changing people's lives for the better and knowing why – and therefore you can keep doing it

- current emphasis on evidence use to evaluate programmes, therefore you need to better understand what it is and how it works.

---

**Usually people start with the planning (covered here in Module 4), but in our experience many people jump in too quickly. We have therefore structured the handbook as follows in order to lay the foundation for effective evaluation:**

## Context

**Module 1** discusses what the deal is with evaluation, and how to get prepared.

## Groundwork

**Modules 2** and **3** describe the things you need to have in place in order to know if you have made a difference.

## Mechanics

**Module 4** discusses how you should go about planning, commissioning and managing evaluation.

## Words matter

Policy makers, programme designers, service providers and evaluators use a lot of common words to mean specific things. Not always the same specific thing! We have included a glossary at the end of the handbook covering a range of terms and how they are often used. What's important is that you, and whoever you are working with, are using them to mean the same thing. To start us off, here are a few common terms that you will find in the handbook and a brief explanation of what we mean by them in this context.

| What we mean | The word we use | Words other people might use for something similar | |
| --- | --- | --- | --- |
| Any set of activities managed together over a period of time that aims to achieve some sort of positive change for a person or group of people you work with. | Programme | Project Initiative Intervention | *(People might also refer to policy, which might also be a mechanism for achieving change.)* |
| The changes (e.g. in attitudes or behaviours) that are likely or achieved as the result of your programme delivery (or social intervention) in the short-, medium- and long-term. | Outcome | Result Impact | *Note: both these terms are used in varying ways to mean slightly different things. Use with care and be clear what you mean.* |
| The extent to which you have achieved the changes you set out to make for the people you work with. These might be changes for individuals or changes to groups or systems, depending on your objectives. | Effectiveness | Success Efficacy | |
| How far you are delivering your programme with the least possible use of resources to the area of greatest need. **OR** How far you are getting the best possible results for the resources that you have. | Efficiency | People are pretty much agreed on this – some people go further and talk about whether this is improving over time or finding new ways to deliver, but we don't go that far. | |
| The long-term, big picture change at a population, society or place level. | Impact | Strategic outcomes | |

**3**

## Risk alert

### Other things that matter

Evaluation can come with risks. For example, not knowing enough about evaluation, not giving enough consideration to ethical requirements and not having a good understanding of cultural context.

Many of these things can be solved using the right skills, good communication and working together.

**With evaluation, the processes you follow are as important as the content that is produced.**

# Module 1

## Getting ready –
## what is evaluation?

Evaluation is something everyone does every day without thinking about it – we review the value, the quality or the effect of activities throughout the day. The process of making these judgements has been formalised into the concept of evaluation of policies, programmes or initiatives.

Evaluation is:

- focused on specific information needs
- judging action against explicit criteria
- predominantly empirical in nature – it involves direct and indirect observation or experience
- using evidence to produce usable recommendations
- feeding information into the decision-making process.

It is useful also to clarify the things evaluation is **not**:

- general research aimed at exploration and expanding the sum of human knowledge
- auditing of public expenditure – auditing is about checking you have done things correctly (and legally); evaluation is more about have you done the right things
- monitoring of interventions – although you absolutely need monitoring to be able to do it, evaluation applies critical thinking to monitoring and other data. The line between the two can be, however, quite blurry when good monitoring is used well
- assessment and selection of individual projects for participation in a public programme (the word is used to describe that, but it's not what we are dealing with here).

### How might you use it?

There are two ways you might want to use evaluation (jargon alert!):

- to know whether things are on track and to learn from what you are finding out (you might find this called *formative evaluation*), or
- to look back and assess how things have worked and what has been achieved (you might find this called *summative evaluation*).

You might be measuring the same things from the two different perspectives.

As an illustration, if a chef tastes soup he has made, he is doing it to see whether he needs to adjust anything (formative), but you, as the customer, taste the soup to see if it was made well/tastes good (summative). To put it another way, formative evaluation is about steering the ship, while summative is about making sure it arrived in the right place. We'll talk more about this later.

4

# Different types of evaluation

There are lots of evaluators, and thus lots of approaches or types of evaluation. Some of these terms are about what is evaluated, and others are about how the evaluation is done. We've already highlighted the importance of not getting hung up on the jargon, but here are a few of the terms you may see being used, either by funders or by evaluators.

Different methods are appropriate for different situations, and to be honest, to some extent methods come in and out of fashion.

| Technical term | What it is |
|---|---|
| Process evaluation | A method of assessing how a programme is being implemented. Process evaluation focuses on the programme's operations, implementation and service delivery. |
| Outcome evaluation | Focuses on the effectiveness of the programme and its outcomes. Bear in mind that outcomes can be produced in the short term, a longer period or in the long term. You can't do an outcome evaluation too early in the life of a programme (because there won't be many!). |
| Economic evaluation | Looks at what resources are being used in a programme and their costs (direct and indirect) compared to outcomes. This is the evaluation that looks at "how much bang for your buck?". |
| Impact evaluation | Assesses programme effectiveness in achieving its ultimate goals. This is going to be done a number of years after the programme is implemented (how many will depend on the nature of the change concerned). |
| *You might also hear about* | |
| Developmental evaluation | This is an evaluation approach that can help develop social change initiatives in complex or uncertain environments. It facilitates real-time, or close to real-time, feedback to programme staff thus facilitating a continuous development loop. It is particularly suited to innovation, radical programme re-design, replication, complex issues or crises. It is not the solution for every situation. |
| Realist evaluation | Realist evaluation asks the question: "What works, for whom, in what respects, to what extent, in what contexts, and how?". In order to answer these questions, realist evaluators aim to identify the underlying generative mechanisms that explain 'how' the outcomes were caused and the influence of context. |
| Utilisation-focused evaluation | Utilisation-focused evaluation is an approach based on the principle that an evaluation should be useful to its intended users. Evaluations should be planned and conducted in ways that enhance the likely utilisation of both the findings and of the process itself to inform decisions and improve performance. |

Your process evaluation may be able to explain why some outcomes were (or weren't) achieved, so although the timing may be different, you need both for the full picture. These are not mutually exclusive.

## Why evaluate?

At the most basic level, evaluation is part of the process of running your programme well – you need to know whether it is working or not, and evaluation gives you a formal framework within which to do that. You can do it for a number of reasons:

- to improve management and implementation

- for accountability – to funders, but also to your other stakeholders and the programme participants

- being able to establish objectively how much your programme positively impacts on your clients' lives, which enables you to deliver quality services that consistently meet the needs of your community

- to plan for the future – regularly reviewing and monitoring the work you do can help you to be more proactive, flexible and strategic when setting work priorities in a changing environment

- because it's a requirement – this is the worst possible reason to do it, and tends to result in something done neither particularly well, nor used particularly effectively, which is a waste of time and resources.

Funder requirements have a tendency to change. A good monitoring system and evaluation strategy will mean you can deal with this with minimum effort, rather than having to change what you do whenever requirements are altered. It also helps meet the situation where you have to report to multiple agencies in the world of patchwork funding.

### Who can evaluate?

You might need some outside help, but once you've worked through a few things with the help of this handbook, you will be able to get it done.
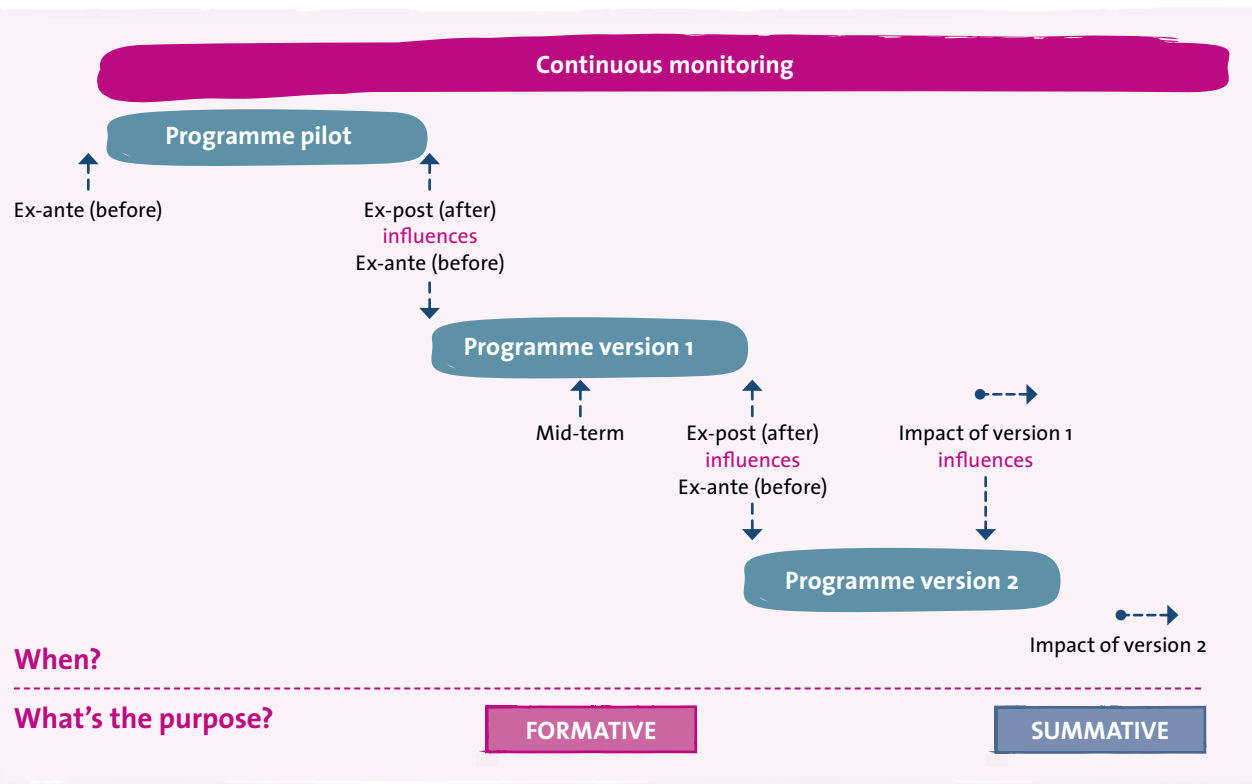
## When do you evaluate and who is it for?

Over the life of your initiative you may want to evaluate a number of times, for different reasons and looking at different things. You might need to provide analysis up-front of the current situation and how you plan to change it (**ex-ante evaluation**, prior appraisal), you might want to keep the programme under **continuous review**, you may want to look back at what you have achieved (**ex-post**). Evaluations feed into each other – the "after" of one phase is the "before" of a new phase. Often you cannot wait for this so you might use information collected at some mid-term point. Remember not all impacts will be apparent until after your initiative has concluded.

Identifying your audience(s) for the evaluation work will help you decide on what to evaluate, at what point. Do you need an evaluation for your funders, to continue your funding? Will that evaluation also satisfy your other stakeholders? Is your evaluation for internal management use so you can improve how your initiative works? You can see that the things you want to know for these are quite different. Your funders might, for example, be interested in how many youth you have helped to improve their job seeking skills. You, as the programme manager, however are as interested in whether the skill building workshops you are providing are working how you thought, and whether they are the best approach.

As we mentioned before, **formative** and **summative** evaluations look at similar things from two different perspectives.

**Continuous monitoring**

**Programme pilot**

Ex-ante (before)          Ex-post (after)
                          influences
                          Ex-ante (before)

**Programme version 1**

Mid-term     Ex-post (after)      Impact of version 1
             influences           influences
             Ex-ante (before)

**Programme version 2**

Impact of version 2

**When?**

**What's the purpose?**

FORMATIVE                                    SUMMATIVE

## Formative

(forward looking, the objective is learning/steering)

• Are you doing what you planned?

• Is it being done well/is it working as expected?

• Are the changes you wanted to see happening, or are there signs of this change beginning?

• If the changes seem to be happening, to what extent are they directly a result of your initiative?

## Summative

(backward looking, the objective is assessing what happened)

• Did you do what was planned? *(How much did you do?)*

• Was it done well?

• Did the planned changes happen/are they on track to happen? *(Is anyone better off?)*

• Are the observed changes attributable to the initiative?

# Getting ready to evaluate

Before you can evaluate, you need to get evaluation ready. There are many tools to help you do that, so here we just cover the key elements. Leaders, managers, staff and volunteers all have a role to play in building the organisation's evaluation capacity. However, the role of leaders and managers is critical in ensuring that everyone in the organisation understands the need and what their part is in the process.

## What you need to sort out

This handbook will take you through what you need in order to evaluate your programme (or portfolio of programmes). However you will need to look at a number of organisational factors, and get a few penguins in a row before you start. This will include your relationships and the governance of your evaluation.

Generally speaking there are few prerequisites for undertaking a *process evaluation* as the focus is on programme delivery and how it operates in practice, and on providing information for potential programme improvement. The two main preconditions are that:

- A planned programme actually exists, meaning that it has been planned as a discrete initiative for a specific purpose, which is separate from, but sits alongside, other practice.

- The programme has been operating for long enough so that you can document and assess its development and operation.

If you are looking to do an *outcome evaluation* there are five areas that you need to be clear on:

- **Programme design.** This includes, for example, that programme outcomes are clearly specified, measurement tools and processes are in place and used consistently, and the way in which change is intended to occur is clearly set out.

- **Communications of your findings.**

- **Programme data.** This includes knowing what data you will need about programme delivery (such as client details) and outcomes measurement, how you will collect and manage client data, and how you will access aggregated data about your clients for analysis and reporting. You might also need some other data for comparison purposes, but if you have identified how you will measure things you should already have this in mind.

- **Organisational context.** This means that the organisation is committed to the evaluation, wants and will use the evaluation, can commit the necessary time and resources, and can engage in the evaluation process such as getting programme participants involved (for surveys or interviews, and so on). It also means ensuring that relevant external stakeholders are comfortable with an evaluation occurring.

- **Cultural context.** This includes application of multi-cultural competencies and concepts, as well as consideration of the principles of Te Tiriti o Waitangi. For example, you should consider appropriate consultation and engagement processes, and also identify how your initiative responds to Māori needs and aspirations. These are often very straightforward, but if in doubt get expert advice.
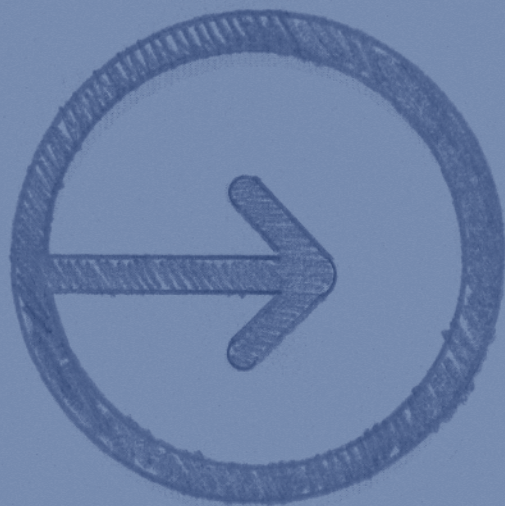
## Useful tools and guides

A toolkit in four parts to help you assess and develop your evaluation capacity
superu.govt.nz/resources/getting-ready-to-do-evaluations

Lessons on evaluative capability and preconditions for undertaking an evaluation – a summary report of two case studies
superu.govt.nz/resources/evaluations

A step-by-step guide to planning and managing an evaluation
betterevaluation.org/en/managers-guide

Evaluation Capacity Development Group Toolkit – a detailed guide
betterevaluation.org/en/resources/toolkit/ECDG_toolkit

## Glossary

A useful list of sometimes interchangeable terms and definitions
superu.govt.nz/resources/glossary

# WARNING:
## The following content contains penguins [1]

They are not just cute; they tell a very important story. We encourage you to listen to them.

Why penguins? Firstly, because in the social sector we deal with people, some of whom are in quite difficult places. We use the penguins to depersonalise the concepts. Secondly, because most, if not all, the people using this handbook will be experts in their field. Sometimes knowing too much about the context makes it difficult to focus on the bigger picture. Once you have understood the concepts via the penguins, it will be much easier to apply them to people, and the real life, complex and messy situations you actually have to deal with.

1   No penguins were harmed in the making of this handbook
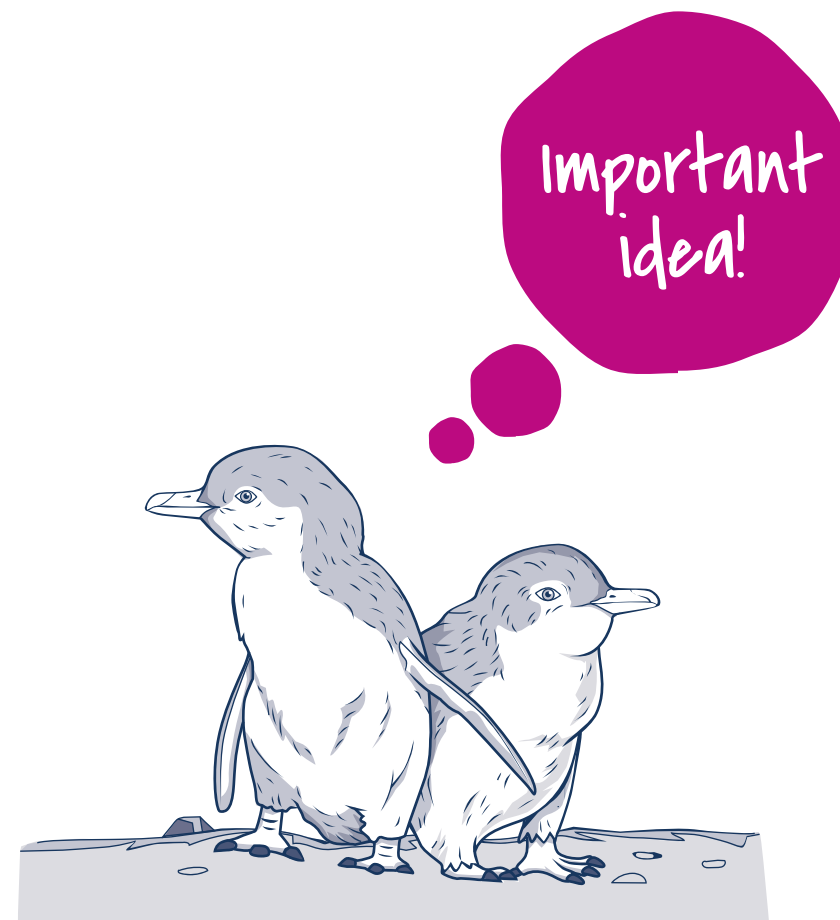
Important idea!

When it comes to people, it is really important to take all your relevant stakeholders on the evaluation journey. You need to have thought about and deliberately planned for how you will include them in the process so that the evaluation helps everyone to learn.

# Module 2

## Understanding your programme

**At the end of this section you should be able to:**

- understand theory of change and logic models and how they can improve your programme

- know how to develop your own logic model, who to involve and how to engage stakeholders

- have an idea of skills, tools and techniques that can help when developing a logic model.

*All these things should be in place before you undertake your evaluation. If you haven't already thought about these things, you need to now.*

## Cutting through the jargon

| What we mean | The word we use | Words other people might use for something similar |
|---|---|---|
| A comprehensive description of how and why a desired change is expected to happen in a particular context. It tries to understand the causal links between what you do and what change occurs – why you believe it will happen. | **Theory of change** | Programme story<br>Programme theory<br>Intervention logic |
| A diagram or picture of your theory of change that shows in simple terms how what you are doing is expected to lead to the changes you intend. | **Logic model**<br>**Intervention logic** | Logical framework<br>Logframe[2]<br>Benefits map (software and business types use this one)<br>Results chain |
| A tool that helps you to link what you want to achieve and how you will measure success. | **Outcomes framework** (see Module 3) | Most people use this term. However what's included in the framework might be broader or narrower according to the situation or how your sector traditionally does this. |

2  Technically speaking a logframe is a table, not a diagram. You can find out more at theguardian.com/global-development-professionals-network/2015/aug/17/how-to-write-a-logframe-a-beginners-guide

# Getting started

In this section we are going to look at how to construct a logic model to help represent your programme, and explain what you want to achieve, how, why you think it will work and how you will know if it does.

You should use a logic model when you are:

- planning a new initiative or programme (including for funding applications)
- implementing a programme
- training staff and informing stakeholders about the programme
- evaluating whether the programme is achieving what you want it to
- advocating for your programme
- improving your programme (e.g. systems and processes).

## Ask yourself a few key questions first

Before trying to set out your objectives, it is wise to be clear what the issue, problem or opportunity is that you are addressing. It is not always the case that what is being treated is the real issue. Of course sometimes this is for good reason, but it may also be because there was a lack of analysis at the start.

Even where the analysis is very good, it helps to set it out logically. This is particularly helpful in making sure that when you think you have agreed on the problem, everyone involved has the same understanding. It will also help to explain the importance of your initiative to potential funders.

So before rushing into things, it's worth taking the time to make sure you are starting in the right place.

An often-used example of addressing the wrong problem is seeking more ambulances to pick up bodies at the bottom of the cliff, rather than improving safety features at the top, which would be a better use of resources and lead to better outcomes.

If you are going to leap into evaluation, make sure you are addressing the right problem. This handbook will help you gain the confidence to swim.

## Clarify the situation (issue/opportunity/ need/problem) you want to address

Many tools exist for defining the issue or opportunity the programme is seeking to address, ranging from very informal to very formal. One tool that is useful in either setting is the construction of a problem tree. This is used to identify the causes, problems and effects in a situation and show how they interconnect. If you are using the logic model process in planning or developing a programme, then effort here will be well spent. If you are using it for evaluation retrospectively, then it is useful to check whether the programme was targeting the right 'thing' (and whether that is still the right 'thing').

For example, here is a very simple problem tree that looks at describing the problem of debt. It identifies how the causes and effects of debt may be related.

You don't need to represent this as a formal diagram like this. There are many examples of different approaches and styles – some more helpful than others. Choose the one that works best for you and your stakeholders.
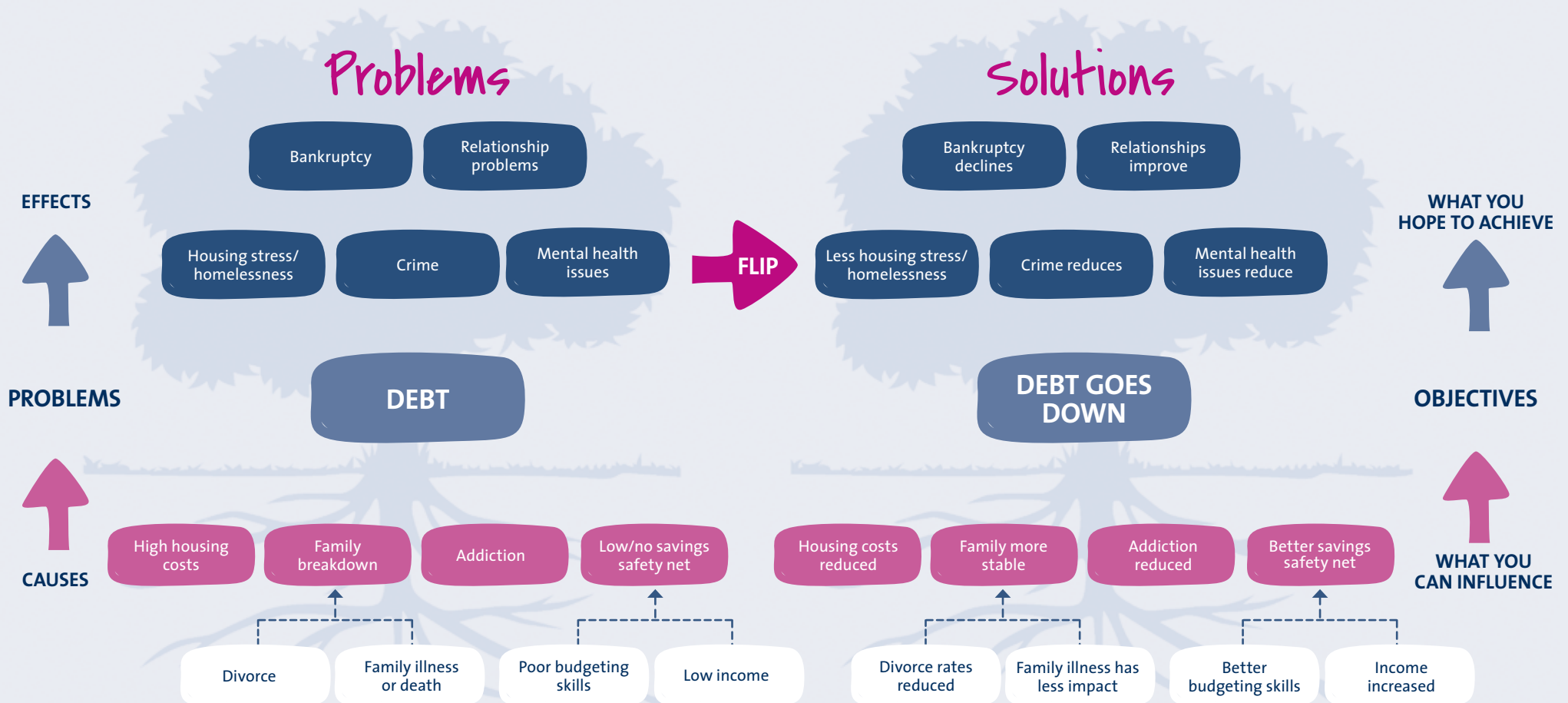
13



**EFFECTS**

**PROBLEMS**

**CAUSES**
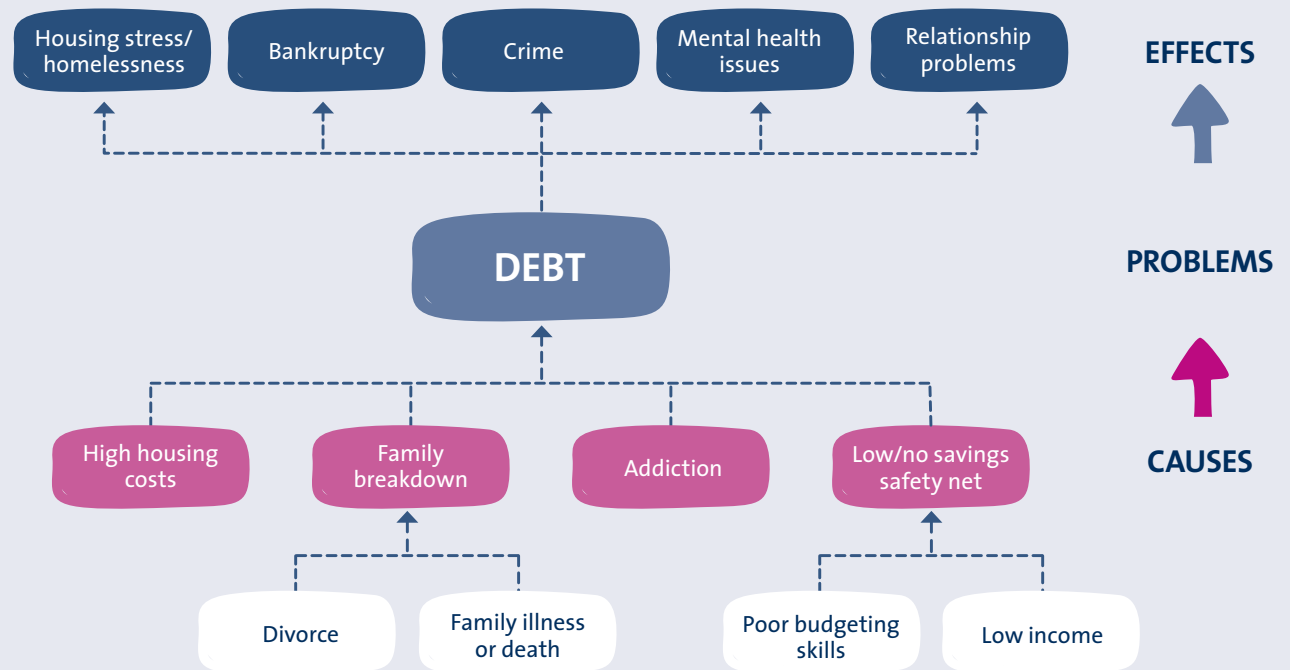
Obviously life is not that simple and there are many links not shown!

## Different representations of the same situation

The point of a problem tree is to:

- make sure you have identified the issue you want to address
- identify your priorities – what are you trying to fix, and what are other things you need to take account of?
- help you set out the changes you want to achieve.

A clever trick, once you've worked out your issues or problems, is to flip your analysis and show it as an objectives or a solutions tree. Debt (the problem) becomes reducing debt (the objective). You may want, once you've drawn the whole picture, to identify where your initiative fits – for example is your role around family support actions, or about financial capability? You might focus on what to do to address those particular issues.

## Problems

| | |
|---|---|
| Bankruptcy | Relationship problems |

**EFFECTS** ↑

| | | |
|---|---|---|
| Housing stress/ homelessness | Crime | Mental health issues |

**PROBLEMS**

**DEBT**

**CAUSES** ↑

| High housing costs | Family breakdown | Addiction | Low/no savings safety net |
|---|---|---|---|

| Divorce | Family illness or death | Poor budgeting skills | Low income |
|---|---|---|---|

→ **FLIP** →

## Solutions

| | |
|---|---|
| Bankruptcy declines | Relationships improve |

**WHAT YOU HOPE TO ACHIEVE** ↑

| | | |
|---|---|---|
| Less housing stress/ homelessness | Crime reduces | Mental health issues reduce |

**OBJECTIVES**

**DEBT GOES DOWN**

**WHAT YOU CAN INFLUENCE** ↑

| Housing costs reduced | Family more stable | Addiction reduced | Better savings safety net |
|---|---|---|---|

| Divorce rates reduced | Family illness has less impact | Better budgeting skills | Income increased |
|---|---|---|---|

14

## Setting objectives and identifying outcomes

Once you have identified what you are trying to achieve and what the change is that you want to make, you now need to look at how to explain the process. This is what's known as a logic model.

The process you should use depends on your organisation and your stakeholders. The best logic models are produced through a collaborative process, involving a range of voices. Although it seems like a lot of resources to involve, a well discussed and agreed model will be much more valuable and useful than one which has been drawn up by one person.

# What is a logic model?

A logic model is a representation – generally a picture – of how your initiative or programme is supposed to work. It explains why what you are going to do is a good solution to the issue you are addressing.

An effective logic model will make a statement about how you expect to bring about change and what you expect to see for your community. It provides a common language and a common point of reference for the people involved in the programme – be they funders, service providers or the people the programme or initiative is designed for or with.

It focuses on explaining the logic behind the process.

## Example of a logic model

You have a headache – that's your issue. You have some resources (pills), you undertake an activity (take the tablets) and you expect to feel better as a result. You can represent that in a picture:

| HEADACHE | PILLS | TAKE PILLS | PILLS TAKEN | FEEL BETTER |

You have some assumptions too – we are making assumptions about what kind of pills they are, for example. If they are not the appropriate medicine, then your logic will break down. So for this process, you have an "if, then" logic:

| HEADACHE | PILLS | TAKE PILLS | PILLS TAKEN | FEEL BETTER |

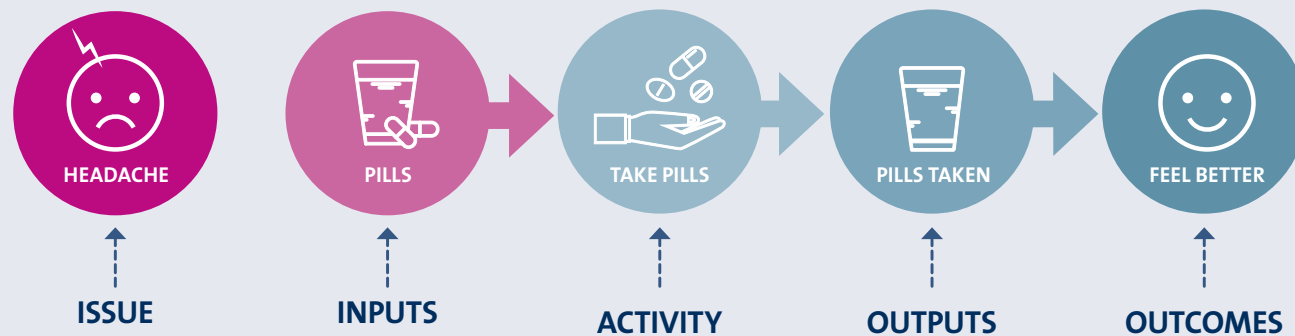**IF**

**THEN SOMETHING HAPPENS**
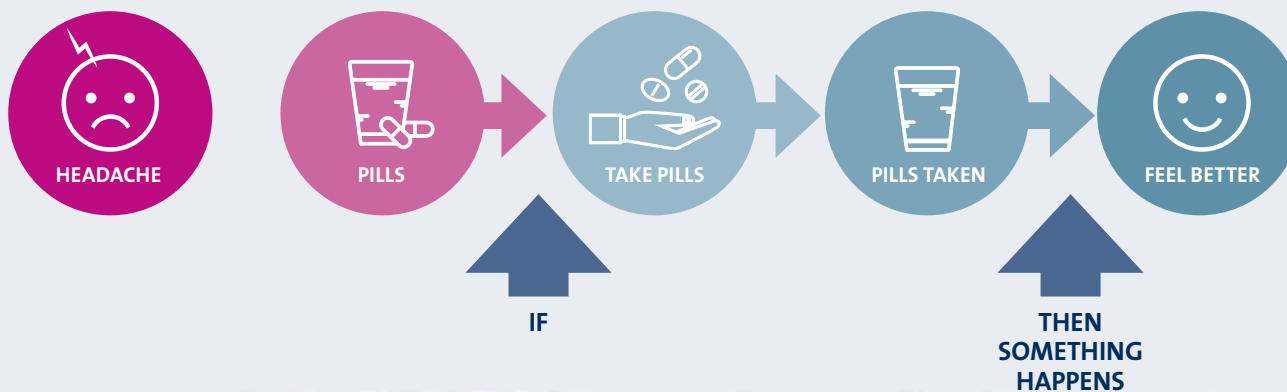
**Top tip**

Spell out all the assumptions being made. You'll need to check them later to see if they still apply. Situations change, knowledge improves – your assumptions may need to change.

Top tip

Outcomes can be short-, medium- or long-term. Make sure everyone involved understands what these timeframes are in your case.

**Example of a logic model** *(continued)*

We can apply labels to the various components of the logic chain:

| ISSUE | INPUTS | ACTIVITY | OUTPUTS | OUTCOMES |
| --- | --- | --- | --- | --- |
| HEADACHE | PILLS | TAKE PILLS | PILLS TAKEN | FEEL BETTER |

**Inputs** are the resources that you bring to the table. Often this is money, but it may also be people, knowledge, facilities, equipment, time, contributions in kind and a range of other things.

The **activity** is what you do: the process that transforms these inputs into something else. Activities might include training, mentoring or social work activities of various kinds. (It could also include, for example, building houses.)

**Outputs** are the thing you produce through your activities. A rough rule of thumb is that outputs should be countable – people mentored, people completing training courses, different types of social work provided, houses built and so forth.

**Outcomes** are the changes that happen as a result of what you did. Depending on the objectives of your programme, you might have a reduction in youth offending as a result of the mentoring programme, people returning to the labour force, improved parenting skills, people in warm, dry houses and so forth. You will need to look at what change you wanted to happen, what change actually happened and when, and what change didn't actually happen (and why).

17

# The Rena -
## an adventure in outcomes

On Wednesday October 5th 2011, the Rena, a container ship on its way to the port of Tauranga, struck Astrolabe Reef/Otaiti. Eight of the containers contained hazardous materials, and the ship carried 1,700 tonnes of heavy fuel oil and 200 tonnes of marine diesel oil. By Sunday a five kilometre oil slick threatened wildlife and the area's rich fishing waters. By Monday oil began washing ashore at Mount Maunganui beach. In all about 350 tonnes were spilled into the ocean, with most of the rest being pumped out into tankers.

As well as a wider impact on the environment, the spill severely affected the bird population in the area. Over the duration of the event, 383 oiled little blue penguins – kororā – were admitted to the wildlife recovery facility. Eighty-nine dead oiled penguins were recovered through the response. Just over one quarter of the admitted birds suffered 90-100% oiling.

**350** TONNES OF OIL WERE SPILLED INTO THE OCEAN

RENA

18

# Case study
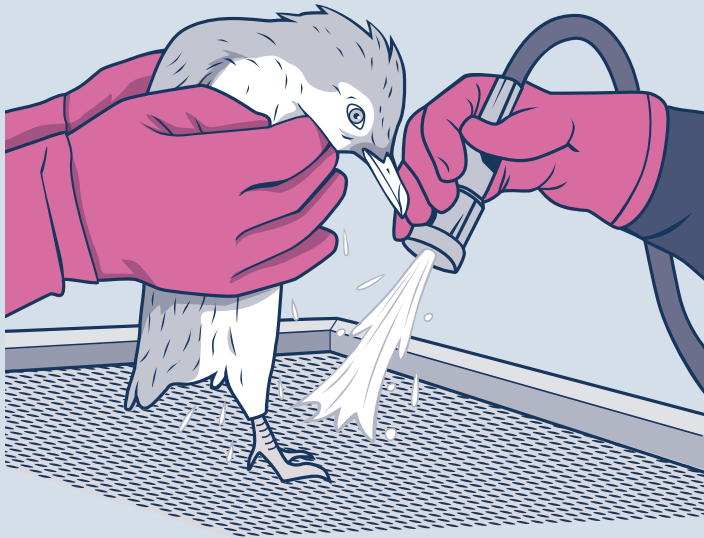
A side effect of cleaning the oil from the birds was the removal of their natural oils, which provide them with protection and enable them to swim without becoming waterlogged. By having their natural oils removed, penguins were no longer protected against the elements. Thus, during the recovery period there was a problem with cold penguins.

A public campaign was mounted to solve this problem by bringing together volunteers to knit jumpers for the penguins to keep them warm until their own oils regenerated.

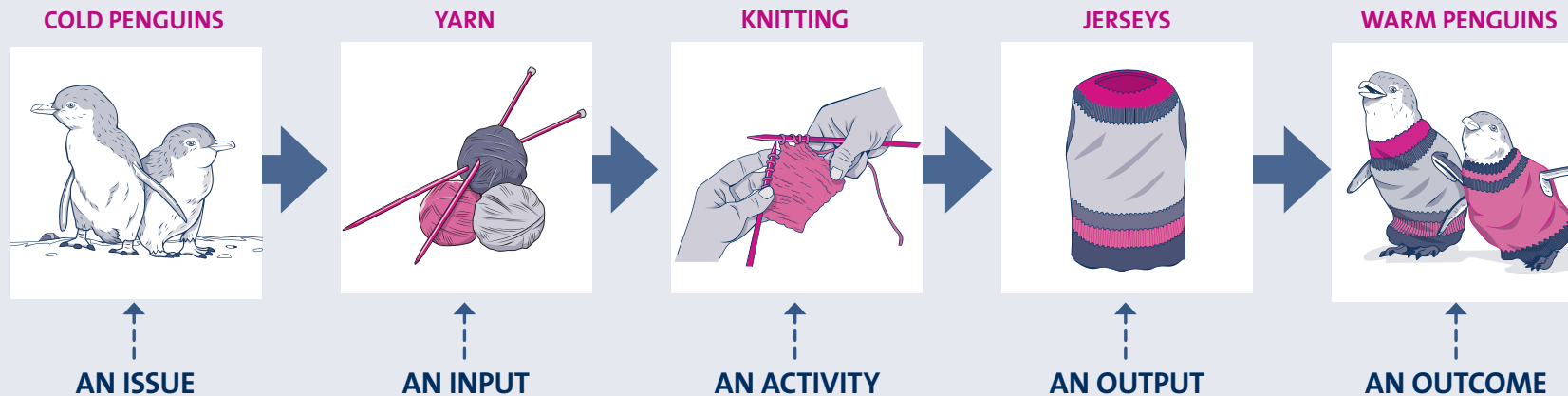**383** OILED LITTLE BLUE PENGUINS WERE ADMITTED TO THE WILDLIFE RECOVERY FACILITY

# Cold penguins' intervention logic

Let's look at how we capture the logic process connecting the problem to the solution. Applying our logic process to this issue we can see that we have all the necessary elements to construct our model:

| COLD PENGUINS | YARN | KNITTING | JERSEYS | WARM PENGUINS |
|---|---|---|---|---|
| **AN ISSUE** | **AN INPUT** | **AN ACTIVITY** | **AN OUTPUT** | **AN OUTCOME** |

This example enables us to be clear about the difference between an output (jerseys) and the outcome, a change you have brought about (warm penguins). It also reflects a number of assumptions, including the key supposition that penguins will tolerate wearing woollen jerseys. In fact, there are strong views to the contrary.

> "Putting that on a penguin's only going to stress it out even more than it already is. These are wild penguins, they haven't had any interaction with humans. There's already enough stress on a bird without trying to put a jersey on it."
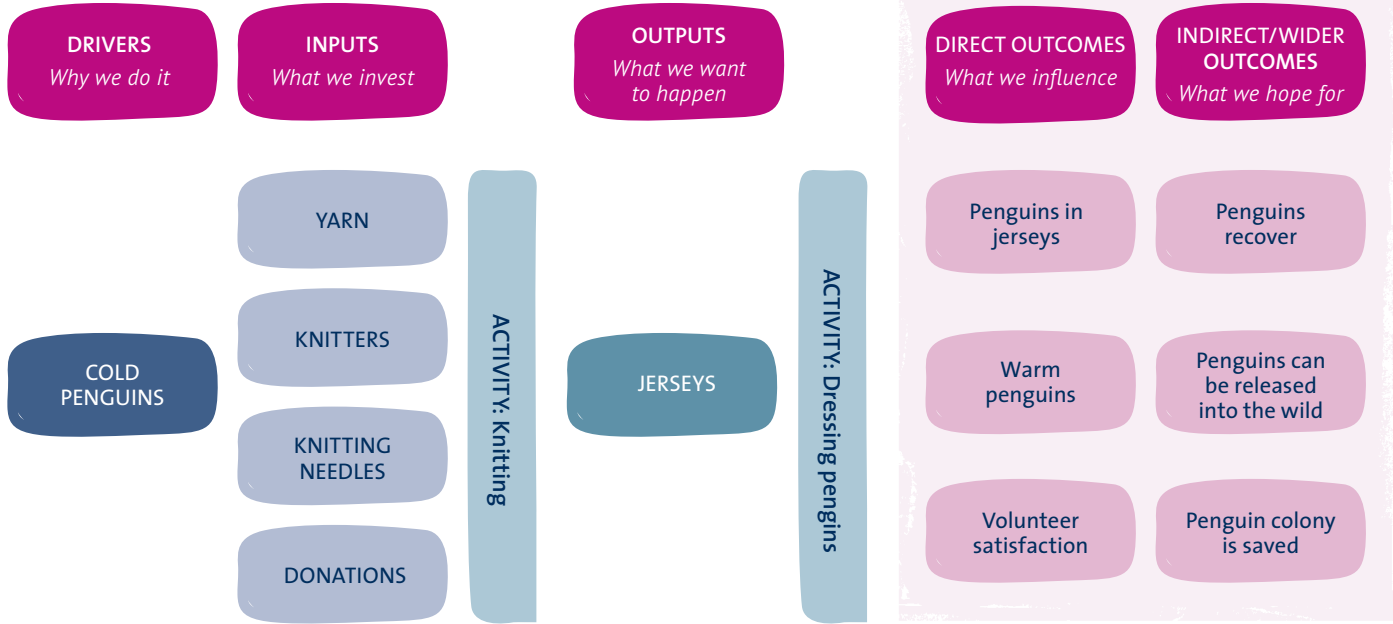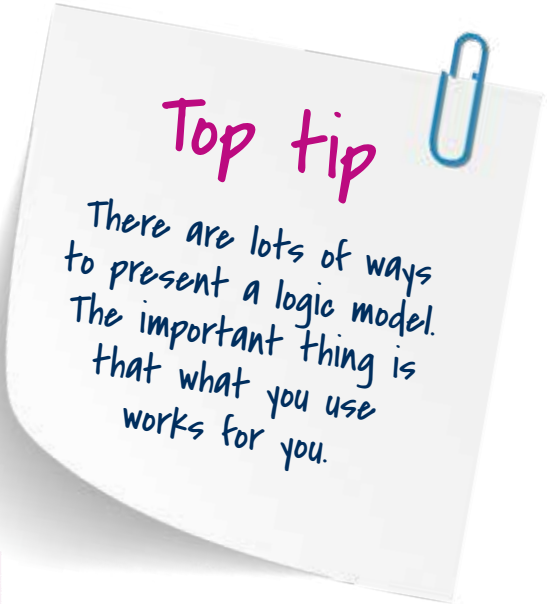>
> *- Natalie Clark, bird/mammal keeper at the Auckland Zoo*

20

## Case study

### Alternative cold penguins' logic model

We can also represent the cold penguins' intervention logic as a slightly more complicated model because we have looked at the bigger picture.

You'll note that we have two sets of activity here – producing the jerseys, and persuading the penguins to wear them. We can adapt the model to show that.

When you are providing a social service you are providing a solution or a set of solutions to a problem or set of problems, so you can take the same approach. Don't forget that many of your outputs will not be something physical like jerseys, but may be things like counselling sessions or advice provided. Another thing to remember in this sector is that sometimes a positive outcome is "something that didn't happen", if your objective is prevention.

### Top tip

There are lots of ways to present a logic model. The important thing is that what you use works for you.

| DRIVERS *Why we do it* | INPUTS *What we invest* | | OUTPUTS *What we want to happen* | | DIRECT OUTCOMES *What we influence* | INDIRECT/WIDER OUTCOMES *What we hope for* |
|---|---|---|---|---|---|---|

COLD PENGUINS

YARN
KNITTERS
KNITTING NEEDLES
DONATIONS

ACTIVITY: Knitting

JERSEYS

ACTIVITY: Dressing penguins

Penguins in jerseys — Penguins recover

Warm penguins — Penguins can be released into the wild

Volunteer satisfaction — Penguin colony is saved

**ECOSYSTEM IS RESTORED**

**ASSUMPTIONS:** penguins tolerate jerseys well; jerseys are the optimal method to warm a penguin

**95%**

of the penguins (365) admitted to the wildlife recovery facility were released back to the wild.

Eighteen admitted birds died (13 required euthanasia and another five died in care).

## A happy ending

On a final note, jerseys have been used successfully to prevent oiled penguins from preening *while waiting to be cleaned* at Phillip Island in Australia. We don't know whether these penguins got stressed. We can say that if the logic was examined earlier in the Rena case, then the jerseys may have been used differently, or not at all. This is a great example of continuing to refine your logic.

The jerseys (and model penguins) are now used as a fundraiser. Buying one helps fund other bird recovery operations. This redirection of resources to achieve a good outcome is a happy ending.

So things may work in some situations and not others. The issue of transferability (i.e. whether it will work elsewhere) might be important for some initiatives. Certainly if you are looking at what works elsewhere, you need to think about this before implementing your own version.
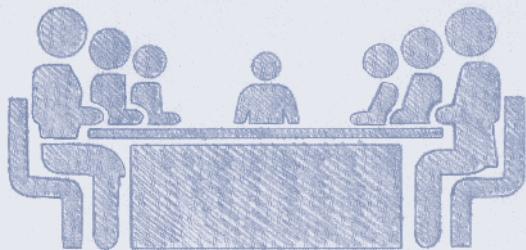
# Top tip

Although your model is just a sketch of your initiative, the more specific you can make it without losing sight of the bigger picture, the easier it will be to use it later.

Bear in mind that not all the viewpoints you need will be at the table, and ensure that these are taken into account.

# How to develop your logic model

There are a number of steps to take:

1. Before you start, work out the scope of your model, taking into account how you want to use it (mainly for planning, mainly for evaluating etc) and who you are developing it for.

2. Find the logic in your existing materials (most people do have all the necessary elements either on paper or in people's heads). You can either set these out in a draft logic model (with gaps if necessary) to use as a starting point or you can just provide them as building blocks for discussion.

3. Bring together a group of stakeholders to construct your logic model, explaining what you intend to do, why and what you expect, and would like to see happen as a result. There are many ways of doing this – whiteboarding, using sticky notes, using tables and so forth. There are also special software programmes that let you do this interactively. Sticky notes work well. They are a low cost option that allow you to move things around as the discussion progresses. It is helpful to have someone who can run the session, is less involved, and can keep things on track.

4. Don't overcomplicate your model. You can make nested or cascading models if you have programmes with several activities or workstreams.

5. Check whether the model makes sense and is complete – revisiting it after a few days and/or having it looked at with fresh eyes can help with this. Make sure you have listed your assumptions and checked that they are realistic. This is an important part of the process.

6. Be ready to revise the model as necessary – situations change and you may need to adapt to meet them.

If your programme or initiative is particularly complex, you may find it simpler to construct one high level logic and a set of sub-models that can go into more detail on different parts or workstreams.

For example, create a main logic and then sub logics for the key outcomes, or a set of logics that looks at things at the strategic, organisational and service levels. This avoids trying to cram too much unnecessary detail into a single diagram.

Note: the diagrams below are to illustrate a point and aren't meant to be read.

**Top tip**

Your logic model needs to be realistic, precise and pragmatic as opposed to aspirational/idealistic. It needs to take into consideration your organisation's constraints (time, resources, staffing etc).
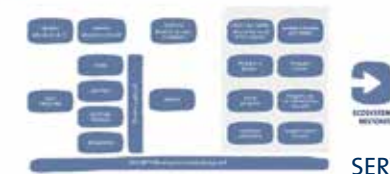
## NESTED MODEL



**OR**
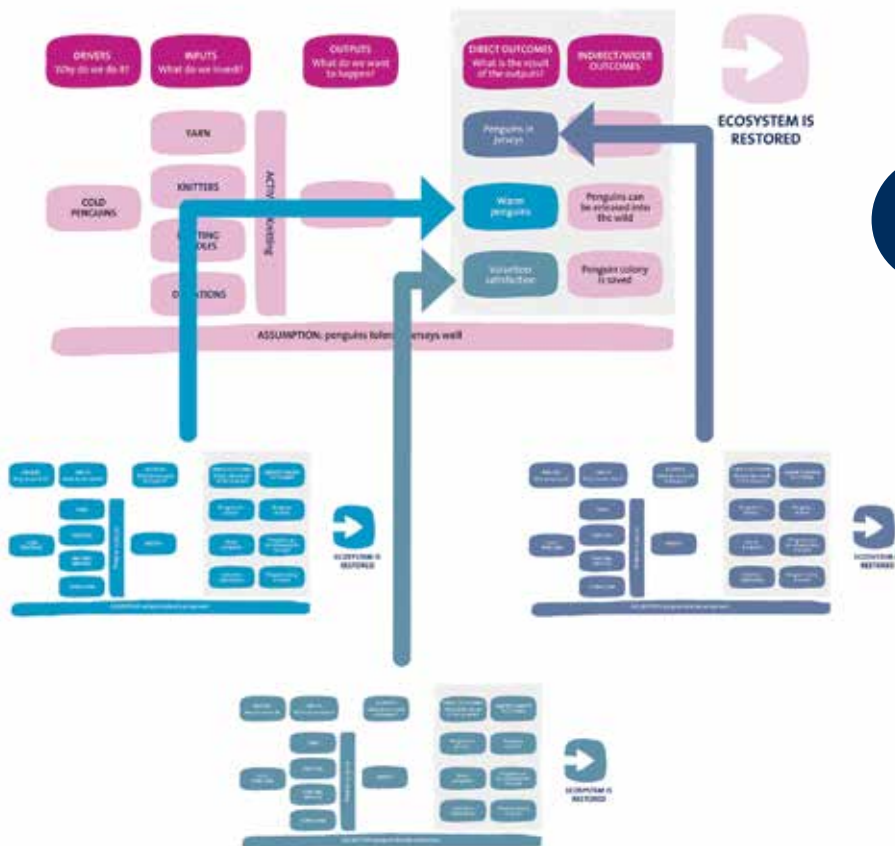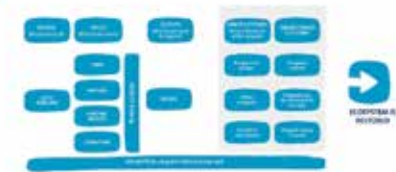
## CASCADING MODEL

STRATEGIC LEVEL



ORGANISATIONAL LEVEL

SERVICE LEVEL

# Further reading and additional resources

If you would like more information about developing and using logic models, there are many sources of support and information. Some of these are relatively generic, and some are more focused on specific issues, for example the field of public health.

## Useful websites

**The United Way "Strengthening Families"**
strengtheningfamilies.unitedway.org/evaluating_steps.cfm

**What Works Aotearoa**
whatworks.org.nz/frameworks-approaches/logic-model/

**University of Wisconsin Extension Services**
fyi.uwex.edu/program-development/logic-models/ has an on-line course, a handbook of the course (200 pages) and a selection of templates and examples

## Guides and books

**Logic Model Development Guide** published by the W.K. Kellogg Foundation and written specifically for non-profit organisations wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide.

**Evaluation Toolkit** from the Office of Premier and Cabinet, NSW Government dpc.nsw.gov.au/programs_and_services/policy_makers_toolkit/evaluation_toolkit

**Logic Model – A Planning and Evaluation Tool** from Public Health Ontario publichealthontario.ca/en/eRepository/Focus_On_Logic_Models_2016.pdf

**Guide 1.2 Developing a Logic Model** by Evaluation Support Scotland evaluationsupportscotland.org.uk/media/uploads/resources/supportguide1.2logicmodelsjul09.pdf

**Logic Models and Programme Planning** from the Centers for Disease Control in the USA cdc.gov/oralhealth/state_programs/pdf/logic_models.pdf

**A 5-minute Guide to Using a Logic Model** published by NESTA as part of the DIY toolkit for social innovation in development (diytoolkit.org) youtube.com/watch?v=6zRre_gB6A4&feature=youtu.be;

**Purposeful Program Theory – Effective Use of Theories of Change and Logic Models** Sue C Funnell and Patricia J Roger ISBN: 978-0470478578 John Wiley & Sons (this is a textbook, and is not free)

# Module 3

## Measuring change

**At the end of this section you will:**

- understand approaches to measuring change

- know why you want to measure things

- know about different things you need to measure.

Whatever the specific objectives of your initiative, the purpose is to bring about some form of change. You also want to demonstrate progress to your participants, funders, programme management and wider stakeholders, and to do that you need objective ways of showing it.

## Cutting through the jargon

You are likely to hear the terms **'measures'** and **'indicators'** used. The difference between them has become a perennial topic of discussion. In many organisations there is a very loose language that merges separate phrases and words such as measures, indicators, performance indicators and key performance indicators (KPIs) into a single thing.

**You need to focus on three questions**:

- what are the things you need to be able to *measure*

- what are the things you need to be able to *judge*, and of these

- which are the most *important*.

For example, a carpenter wants to know that a table is six feet by three feet. **That is a measure.** There is a standard foot out there and we can all agree how long it is. A restaurant owner wants to know that a table will seat six people. **That is an indicator.** It indicates its size, but does NOT measure it (by a universally agreed Metric or Imperial or Troy or any other system).

Whether the measure or the indicator is more useful will depend on who you are and why you need to know this. "A table that seats six" is not detailed enough for the carpenter, and "a six feet by three feet table" requires further processing and a judgement from the restaurateur, so is not immediately helpful to him either. As ever, just be clear among your team and stakeholders how you are using the words.

In this handbook because we are mainly talking about outcomes, we will be using the word **indicator**. Generally we will be talking about measuring or counting something that gives us an *indication* of whether the desired change is happening or likely to happen.

| What we mean | The word we use | Words other people might use for something similar |
|---|---|---|
| Something you can measure or demonstrate that reasonably shows a change that you have influenced | Indicator | Measure<br>Key Performance Indicator (KPI)<br>Metric |
| Something you can measure or count that is the direct result of your activity | Measure | Performance Indicator<br>Key Performance Indicator (KPI)<br>Metric<br>Performance measure |
| The **most important** of a list of potential measures or indicators, that can give you the key elements of what you need to know about your initiative | Key Performance Indicator (KPI) | Key Performance Measure |

# Why use indicators?

Having a good programme or initiative is not enough. You will want to know that it is working the way you thought it would. Within that, different people will have different priorities.

People who use indicators have different roles. They may be service providers, funders, policy makers or even regulators,[3] and they use indicators for different purposes. This might be for general management and planning purposes or for a better understanding of the longer term changes you are contributing to. If you have funders (or are funders) you want to know whether your investment is resulting in the change you want to achieve. Policy makers need to know whether their objectives are being met.

In many cases you can identify ways to assess qualitative change. You cannot easily measure happiness or wellbeing, for example, but you can count the number of people who report an increase, and you can ask them to identify how much it has increased by. There are also a number of tools that look at the components of wellbeing or happiness.[4]

---

3 For example the Charities Commission

4 For example, the Whānau Rangatiratanga Framework (superu.govt.nz/whanau-rangatiratanga-frameworks) or the OECD Framework for measuring well-being and progress (oecd.org/statistics/measuring-well-beingand-progress.htm). For individual wellbeing there are scales such as the Kessler Psychological Distress Scale (K10) framework. Some of these are quite sector specific.

# Scale is important

You can also measure things at different levels, which look at different issues. A small local programme on its own is not going to make a measurable difference to the whole population, nor should it claim to – although it will contribute, the effect on the population will not be measurable. Equally, policy makers are concerned about making a difference to the country, so the effect on a single individual is of less interest than the overall change that is being brought about. Put simply, the programme cares about the outcome for an individual "Casey", while the policy makers are interested in reducing the number of "Caseys" who need help. Sometimes, to know whether your programmes are succeeding, you need to add up the effects on the individual "Caseys". The level is important because it defines what sort of indicators you need and who you are measuring – everyone, a group or a person?

| Level | Questions? | Who wants to know? | What type of Information? |
|---|---|---|---|
| **Population**  | Are you making a difference to the whole population? | Policy makers, ministries, ministers, government agencies, regional or local government | Aggregate data on whole population groups, mainly from administrative data or statistical series |
| **Targeted group**  | Are you making a difference to the group of people you are trying to help? | Funders, programme managers, service providers, policy makers | Aggregate data on the people the service is provided to, mainly from programme monitoring information and evaluation follow-up |
| **Individual clients/families/whānau**  | Are you making a difference to the person or whānau you are trying to help? | Programme managers, service providers, clients | Individual information on the person or whānau the service is provided to, mainly from monitoring information and follow-up |

You also need to know that it was you that contributed to the good outcome for "Casey" – it wasn't just something that would have happened anyway. This is called **attribution**, and we'll look into that next.

## How do you know you made a difference?

Understanding how much of any change was a result of what you did is called **attribution.** It is really important. The critical question is how can you be sure the effect is related to what you did? There are a number of reasons why it might not be. These are outlined in simple terms below. Evaluators may use technical terms for these. We cover these in the glossary online.
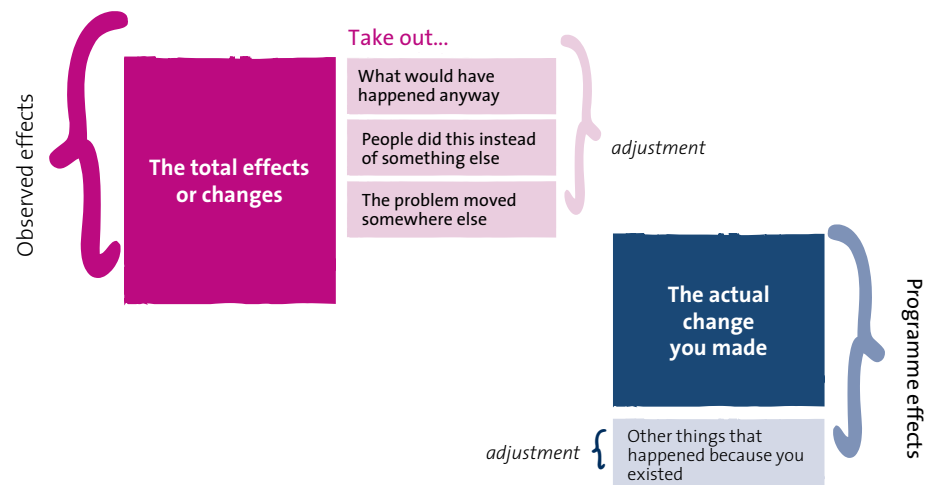


All of these are notoriously difficult to measure, especially in the social sector. However, there are ways of making estimates of some of these – particularly what would have happened anyway.

It makes sense to look at these issues and assess the extent to which you can say something about them. It will be less difficult if you decide up front how you are going to tackle them, as you can then collect the relevant information, where it exists or can be created. It may, for example, be possible to set up a control group of people who do not receive your service to see the difference between them and your clients. If this is not possible or ethical, you may be able to track a matched comparison group through data collected elsewhere. NB if you are not going to do either of these, you should explain why. This is all part of the discussion about what to measure, how and when.

29

## Example

Let's suppose your initiative is a new youth drop-in centre providing assistance to those seeking employment. After a year you have worked with 100 (to make the maths easy) clients. Of these, **75 report that their job seeking skills have improved significantly following your support.** That's a great success rate. However, you know from experience that about five of those would probably have improved these without your intervention, and about 10 of them were previously using another support service a few miles away, so are not "new" clients in the system – they have just substituted your centre for their previous support. **So in fact, you should say that you have probably been directly instrumental in the improvement for 62 people.**

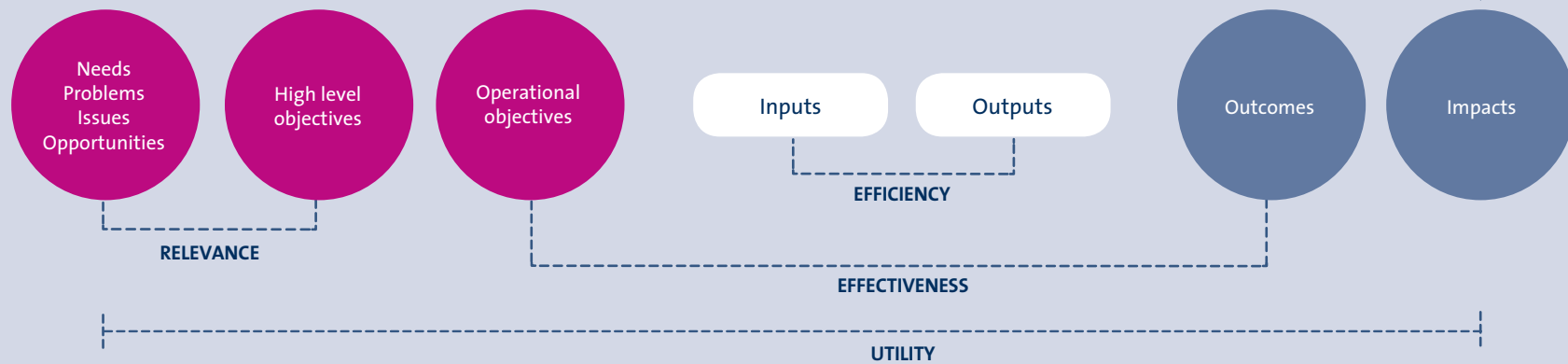| | |
|---|---|
| **Total number of clients** | **100** |
| **Number reporting an improvement** | **75** |
| **HOWEVER** Ten of those were already getting help elsewhere, so would still have the same result if you didn't exist. **Assuming the same success rate (75%), eight (because you can't have 7.5 people!)** would have improved without you, so while you count them as a success for management purposes, to work out the extra difference you made (your impact), they need to be subtracted from the total. | **67** |
| **AND** Subtract the **five who would improved anyway** even if you didn't exist (assuming they were all in the "new client" group – since we don't actually know) | **62** |

**But wait, there's more!** Since your project started there has been a reduction in petty crime and vandalism in the area. It may be reasonable to say that at least some of that could be indirectly due to your initiative. You should get some recognition for that too, although that will be very difficult if not impossible to put a figure on. Nevertheless, it is worth mentioning as an additional benefit to which you have contributed if you can show why this is the case.

## Impacts

is one of the most ambiguous words in evaluation. Always check how it's being used.

# What do you want to measure?

Before setting off, there are different types of evaluative criteria you need to consider for your evaluation plan:

Needs Problems Issues Opportunities • High level objectives • Operational objectives

**RELEVANCE**

Inputs • Outputs

**EFFICIENCY**

Outcomes • Impacts

**EFFECTIVENESS**

**UTILITY**

- For management purposes, you will want to know whether you are using your inputs in the most efficient way to maximise your outputs. This is known as **efficiency**. For example, are you reaching the most people possible with the resources you have? Could you reach more people if you did things differently? Could you reach the same number of people with fewer resources if you changed your delivery approach?

- You and your funders will want to know if you are making the difference you wanted to make – are your outcomes in line with what you wanted to achieve? This is known as **effectiveness**.

- A question sometimes not addressed is **relevance**, that is whether your high level objectives match your identified needs – you may have been highly relevant when the programme started, but a few years down the road the objectives may be less

relevant. Possibly the situation has changed over time, or your programme may have been successful in remedying the situation. So the question might be are we doing the right thing? Or it might be are we still doing the right thing?, depending on timing.

- In the long term, the question should be asked whether the initiative's impact actually ties back to the high level objectives. This is referred to in some circles as **utility**.

- If you are implementing a pilot programme, other things to consider are **scalability** and the **transferability** of your approach to other locations.

Once you are clear on what you want, you can start to assess what you actually need. Something you want might be nice to know but not essential. Other things might not be obvious but are essential elements of something you really need.

# What do you actually need to know?

Let's refer back to our penguins to work this out in practice.

One of our objectives was to have warm penguins. We can ask about the effectiveness of this (are we warming the penguins?) and also the efficiency (are we warming enough penguins with the resources we have?). It is helpful to put that into a simple framework.

-------------------------------------------------------------------------------

**We could ask many questions.**

- How cold was the penguin? (baseline)
- How warm is the penguin now?
- How warm should the penguin be?
- How long did the penguin take to get warm?
- What's the outdoor temperature?
- How many penguins did you save?
- How warm would it have been without the jersey?
- How much is a warm penguin worth?
- How does that compare with the cost of warming the penguin?
- What is the value of penguins?

**How warm is your penguin?**

Which questions are the most important will depend on what you want to know. So, if you want to know the economic (or environmental) impact of saving the penguins, questions around the value of a warm penguin will be important. If your concern is whether your programme is actually leading to warm penguins, then the more immediate question of knowing whether the penguins are warm is of concern to you.

There are different types of indicators that cover different aspects of understanding an initiative and its effects to fully make sense of things. You generally need to use combinations of ways of measuring to fully make sense of things. For example, to know the scale of change, you need to know both where you started from and where you are now.

To understand if the recovery programme worked well, you need to know how many penguins survived, how many wore jerseys, and whether the ones that recovered wore jerseys.

**Top tips**

Just because you can count something does not mean that you should. And just because you cannot count something does not mean you should ignore it.

# Examples of indicators

| What are we looking to measure? | Type of indicator | Example |
|---|---|---|
| What was the situation at the beginning of the process – where are you starting from? You need to know this to identify change. | Baseline | Temperature of penguin once cleaned |
| How much have you done? You need this information for management reporting, and to work out how efficiently you are carrying out your initiative. | Output | Penguin jerseys Penguins saved |
| How much is this all costing? How do the costs break down? This is needed for good management, to report to funders and to work out efficiency. | Financial | Cost of warmed penguin |
| Whether the processes are working well. Measuring efficiency will help you understand some things about your process, for example, waiting times, bottlenecks, etc. | Process | Time to get penguin warm |
| How big is the change that you have achieved? Are you on track to achieve your longer term objectives? | Outcome | Change in penguin temperature Penguin colony saved |
| What factors outside your control might have a significant influence on your outcomes? You might be able to identify some of these, but some may come out of the blue. | Contextual | External temperature |
| What was likely to have happened if you had done nothing. This might be through comparing your clients to a group of people who are not part of your initiative or by other forms of comparative analysis. It is challenging to do and sometimes has moral/ethical implications, but needs addressing. | Counterfactual | Temperature of untreated penguin |
| How much of any change is a result of what you did? Much as you might like to claim all the credit, it is unlikely that you are the sole factor leading to success. | Attribution | Number of penguins that survive **compared to** the number that survive without jerseys (so what proportion of your number is extra penguins saved thanks to you?) |
| The long term, big picture change at a population, society or place level. | Impact | Effect on local ecosystem |
| The value of what has been achieved in relation to its cost. Both the cost and the value may be more than just financial costs or benefits. Crudely put, how much "bang for your buck"? Funders choosing between options will generally pick the one with the higher return. | Economic impact/ Return on investment | Value of warm penguin/ cost of warming |

# Planning to measure

Using a simple table like the one below will help you work out if indicators are feasible and worthwhile. This helps you work out whether you are measuring something that you strongly influence rather than something that could be completely derailed by an outside influence. It also helps work out how you would get the information, when, and at what cost. Knowing these things helps to decide how feasible it is to use that measure. A perfect measure that costs more than the programme budget to collect is not helpful. Neither is something that gives you your answer too late to be useful. **A strong helping of common sense is very useful here.**

| Intended outcome | Success criteria | Programme factors affecting success | Non-programme factors affecting success | How will you measure this? (What indicator will you use?) | Type of indicator | Source of data | Method for collecting indicator | When? |
|---|---|---|---|---|---|---|---|---|
| Penguins recover | Penguins survive | Treatment of penguins (jerseys, etc) | Disease Amount of oil ingested by penguin | Proportion of penguins that survive | Quantitative (summative use) | Count surviving penguins | Observation | After one month |
| Warm penguins | Penguins do not get hypothermia | Treatment of penguins (jerseys, etc) Availability of jerseys | External temperature | Penguin core temperature | Quantitative (formative use) | Penguin temperature monitoring | Testing (all or some penguins) | Daily |
| Penguins can be released into the wild | Penguins thrive | Conditions at rescue centre | The nature of penguins | Penguin behaviour is normal | Qualitative | Observation Research evidence on penguin behaviour | Observe penguins | Weekly |

**YOU CONTROL**   **YOU DON'T CONTROL**

33

What you are looking for is a small number of perfect indicators. Obviously that's unlikely to be possible, but a large number of only vaguely-helpful-at-best indicators is almost as unhelpful as too few (and an expensive waste of resources). It is therefore worthwhile spending time looking at your intended outcomes and asking:

- How will I know if I have achieved that (what does success look like?)

- How much control do I actually have over that?

- How can I measure that? Will it be numbers (quantitative) or will it be more qualitative?

- Where can I get the data from, and how will I collect it? Will this be something we ask participants and stakeholders, or will we use data that has already been collected?

- When will the data be collected or available?

Reviewing all that, ask yourself whether it is a sensible way to try and understand whether you are on track to achieve your outcome? If not, what else might you do instead? This helps keep things on track (only measure what you need!). The review also indicates **who** should be collecting the data (is it programme or is it agency, for example).

## Building an outcomes framework

You can use the same structure as in your logic model to draw up a framework for your outcomes, to derive your indicators and bring this together into an outcomes framework. This extract from a real example illustrates this.

**WHAT YOU WANT TO ACHIEVE** ←------------------------------------------ **HOW YOU WILL MEASURE SUCCESS** →

| High level outcome | Key outcomes | Potential measures |
|---|---|---|
| Our whānau live in safe, stable homes | Whānau live in quality homes | Rates of housing-related acute hospital admissions (asthma, respiratory diseases, rheumatic fever) |
| | | Neighbourhood survey measure e.g. using the same questions as the General Social Survey (GSS) NZ, on condition of house and incidence of dampness and mould to allow comparison |
| | | Tenancy survey measure (social housing tenants only) |
| | Whānau enjoy housing security | Proportion of income spent on housing (relative to national or regional average) |
| | | Homelessness rate (if available at local level) |
| | | Number / proportion of people (A and B priority applicants) on the Social Housing Register |
| | | Number of people receiving Emergency Housing assistance or Accommodation Supplement (two indicators) |
| | | Length of tenure (Social Housing) |
| | | Length of tenure (Census) |
| | | Number / proportion of home owners |
| | | Housing Affordability Index (if available at relevant level) |

> You do need to distinguish between wanting to know (it's interesting, but not completely relevant) and needing to know (it's important for my programme/funding/policymaking etc). Your indicators should be firmly tied back to your questions and they should make sense.

Remember that just as you have taken timing into account in setting out your outcomes in your logic model, the same will be true of your indicators. There will be some things you can measure quite quickly, but others will not be noticeable until further down the road. For example, if your objective is to reduce reoffending, and your measure is the number of people who do not reoffend within a 12 month period, you can't actually produce that information until after the 12 months is up. You **may** (depending on where you get your data from) be able to see if you are on track before that, but you won't **know**.

## Problems with indicators

The most common problems we see with indicators are:

- An absence of indicators (particularly at the level of impacts) – data is just not collected or is not used.
- An excessive number of indicators – data is measured out of convenience regardless of whether it's useful.
- Badly defined indicators – indicators that lump several things together that might not be related, or indicators that might mean different things to different people, so the data is not collected consistently, but changes according to who did it.
- Indicators that are strongly influenced by the context and weakly influenced by the programme. These won't tell you anything about what you have achieved as external factors will have more influence than you.
- Indicators that are difficult/expensive to quantify – there are many sophisticated things you **could** do. Whether you should is a question you need to reflect on in the light of your programme, your needs, and your budget.
- Indicators that are inadequately presented and explained – if you have gone to the trouble to be able to measure something, make sure you explain what it shows so that your stakeholders understand it. A pie chart is not necessarily the most informative way to present data, for example.
- Indicators that induce perverse behaviour – it's human nature to "manage to the measures". Make sure your measurement system doesn't make it easier for people only to treat the easy cases, for example. Or to manipulate waiting times by changing definitions.

There's something about indicators that appears to lead rational people into all sorts of strange decisions.

# Further reading and additional resources

**Making Measures Work for You:** A guide for non-profits on how to use indicators effectively. The guide looks at tensions that drive the debate about outcomes measurement, as well as common questions about its potential risks and rewards.
grantcraft.org/guides/making-measures-work-for-you

**The Evaluation Framework for Peer-Based Youth Programs:** This user guide provides instructions designed to help youth service providers develop an evaluation plan to measure the effectiveness of their programmes.
mypeer.org.au/program-planning/evaluation-framework/

**Performance Indicators for Evaluating the Mental Health System:** A framework for evaluating the mental health system in California. This is a full set of measures at the different levels (from individual to system) that shows how the whole picture can be built up.
dhcs.ca.gov/services/MH/Documents/03%20Performance%20Indicator%20
Proposal%20final.pdf

**The RMA Quality Planning Resource (NZ)** gives a clear explanation of indicators that reinforces this module.
qualityplanning.org.nz/index.php/monitor/tools-and-indicators

**Indicators:** The book chapter from the *Search for Common Ground* goes into more detail than this handbook and includes an indicator quality checklist.
sfcg.org/Documents/dmechapter4.pdf

**Health Compass**, from the Health Communication Capacity Collaborative (sponsored by USAID), has a clear guide on developing measures with glossaries and worksheets.
thehealthcompass.org/how-to-guides/how-develop-indicators

**Selecting Performance Indicators:** Four pages of performance monitoring and evaluation tips from USAID.
pdf.usaid.gov/pdf_docs/PNABY214.pdf)

# Module 4

## Evaluation planning and management

**At the end of this section you should have a better understanding of:**

- what you need in an evaluation plan
- how to identify your key evaluation questions
- how to design and manage your evaluation, and where to go for help
- assessing the quality of your evaluation
- ensuring evaluation findings are well communicated and used.

## Be an early bird

We know that starting your evaluation process at the end is too late – the opportunity to collect some data will have been missed and your findings risk being incomplete and/or too late to be useful. The most useful evaluations occur when a programme has been designed with evaluation in mind. To get things done at the right time with the minimum of effort, you need a plan.

Your monitoring and evaluation plan sets up the necessary systems and processes to collect and analyse the data and information needed to optimise programme performance. Evaluation should generate accurate, objective and evidence-based information to help managers make sound management decisions and demonstrate success. It also provides the feedback to check that the programme continues to be relevant and is being implemented in the most cost-effective way. Good (useful) evaluations can only be carried out if a programme has been designed with evaluation in mind – it is an integral part of the steering process.

The programme plan should set out the broad outlines of how the programme will be evaluated. A more detailed evaluation plan should be a separate document produced when you actually start the evaluation process.

Your plan should identify how you will collect and report performance data on a regular basis and act on the performance indicators. You can't evaluate a programme without good evidence.

## Who develops the evaluation plan?

It's helpful if the plan can be developed by someone with research or evaluation experience, but it must have inputs from programme staff involved in designing and implementing the programme. There should be buy-in from key stakeholders.

## What goes in the plan?

The main things you need to cover in the plan are:

- what needs to be evaluated (an outline of the programme)
- the purpose and scope of the evaluation
- your key evaluation questions
- your timeline
- available resources
- risk management
- your outcomes framework and associated indicators.

Don't get carried away – this is a plan, not an encyclopaedia – if you make it too detailed, not only will it be out of date very quickly but nobody will use it. It is a living document that should be referred to and updated on a regular basis.

## What are the key evaluation questions?

These are the high-level questions that your evaluation is designed to answer. They are **not** specific questions that are asked in an interview or a questionnaire. An agreed set of questions makes it easier to decide what data to collect, how to analyse it, and how to report it. You should not have many – five or six high-level questions is more than enough. Each of them will have sub-questions that you need to answer in order to address the high-level questions. Try to focus on questions you have a fighting chance of getting an answer to – it might be nice to have insights on deeply significant philosophical questions, but in reality it won't happen so there is no point in putting resources into it.

**If you are evaluating the outcomes of the programme, you will be asking questions such as:**

- How well did the programme work?
- Did the programme produce or contribute to the intended outcomes in the short-, medium- and long-term?
- For whom, in what ways and in what circumstances were outcomes produced?
- What unintended outcomes (positive and negative) were produced?
- To what extent can changes be attributed to the programme?
- What were the particular features of the programme and context that made a difference?
- What was the influence of other factors?
- What innovations does the programme present?

**If you are evaluating how the programme is working, you will be asking questions such as:**

- How is the programme being implemented?
- How appropriate are the processes compared with quality standards?
- Is the programme being implemented correctly?
- Are participants being reached as intended?
- How satisfied are different groups of clients?
- What has been done in an innovative way?
- What, if any, things have happened that were unexpected or affect other people?

**Here's an example of key evaluation questions from a place-based initiative:**

| What outcomes did the programme achieve? | Was the collective impact approach effective? | Did the programme operate as intended? | What can be learned and transferred to other settings? |
|---|---|---|---|
| What effect did it have on the residents? | How did the organisation function? | Was the programme designed to succeed? | What worked, what didn't and why? |
| What effect did it have on housing? | To what extent did collective working occur? Did the community have an effective voice? | Have provisions been made to continue activities beyond the current funding round? | How much of the programme is transferable and what relies on local conditions? |
| What effect did it have on community development? | To what extent did collective working lead to adaptive or innovative service delivery? | | |
| What effect did it have on service provision? | To what extent was the community strengthened? | | |

There are lots of methods you can use to answer these types of questions. This is where an expert evaluator comes in. Make sure you have thought about everything up to now before you rush in and apply any method.

**!**

**Ethics alert**

Ethics are important when dealing with penguins. They are far more important when dealing with people. You need to address ethical considerations in your evaluation plan.

Let's just do a survey!

No, let's do a survey when we know it is the right tool and we've planned this properly!

# Budgeting

The question everyone asks is "how much should I spend on an evaluation?" There is no right answer – the costs of evaluation should always be reasonable, in relation to the programme itself – **the aim is to improve implementation not to divert resources from the programme and its clients.** That being said, there are a number of factors that can help you make a judgement:

- **Is this a pilot programme?** If so, it's worth putting a bit more resource into evaluation as the purpose is to learn what works, how, when and why.

- **Is this a well-established programme that's existed for many years and has had a number of evaluations over that time?** If so, a light touch may be appropriate at this point, as there is already a body of knowledge.

- **Is this a "high risk" programme?** What a funder thinks is high risk is not necessarily the same as a provider, but factors that can put it in this category could include the profile of the initiative, the level of funding, the target audience/client group, level of innovation or random political factors. If it falls in this category, it is likely to have a higher degree of scrutiny so would justify more investment in evaluation quality and coverage.

- **What is the scale of the initiative?** If this is relatively simple and small scale, a small budget is all that you will have. If it is large scale and/or complex you will need more resources to evaluate effectively.

- **Are you going to evaluate internally or externally?** If done internally it may feel like it costs less as things can be absorbed in the staffing budget. This may be an illusion, however. If you are using an external evaluator, then you will need to make provision for their fees. You will also need to budget for time to manage the process.

# Commissioning and managing an evaluation

One of the most important parts of commissioning and managing an evaluation is deciding who is the client, and what will be the process for signing off the report. This has to be set out before you start, or you risk going round in circles trying to get final approval of the report. If you are using outside evaluators either to design or to carry out the work, be very clear on what you want to know, and heed their advice on how to do it. It is sadly not uncommon to see evaluation briefs that are very detailed on how ("we want a survey") without actually saying what the real question is.

## Internal or external?

Depending on the timing, nature and complexity of an evaluation, it may be carried out internally or by external experts. There are pros and cons to both methods.

| | Strengthen independence and objectivity | Support organisational learning | Optimise use of scarce resources | Mobilise evaluation expertise | Mobilise in-house subject expertise |
|---|---|---|---|---|---|
| **External evaluation** | 🐧 | | 🐧 | 🐧 | |
| **Internal evaluation** | | 🐧 | | | 🐧 |

If the main focus is **accountability**, then the use of an external evaluator can help demonstrate independence and objectivity. Remember, they will not have the detailed knowledge that might be expected from an internal evaluator – this does have some resource implications as they bring themselves up to speed.

If the main focus is **organisational learning**, then this is easier when the work is integrated into your internal activity. If you have no resources internally then your options are limited.

> Increasingly people are moving, especially for process evaluations, to a collaborative model where all sides can bring their experience to the table. There is no correct approach, only the one that fits what you and your stakeholders need best, within the resources that you have.

## Finding an evaluator

At the moment there is no formal certification of evaluators in New Zealand (or anywhere except Canada), although there is a lot of discussion about whether and how there should be globally. There are, however, professional organisations where you can find listings of experienced evaluators.

In New Zealand we have both ANZEA, the Aotearoa New Zealand Evaluation Association (anzea.org.nz) which has a directory of professional evaluators, and the Australasian Evaluation Association (aes.asn.au) which also has a consultants' directory. Both organisations allow you to list tenders on their website and have codes of behaviour to which people listed in their directories must subscribe.

## Designing a terms of reference (TOR)

Your terms of reference for the evaluation are different from the plan – the plan should cover all the evaluation you envisage doing, whereas the terms of reference are for a specific evaluation exercise. All or part of the TOR may form part of a Request for Proposals (RFP) if you are using an external evaluator. You need to set out your expectations and requirements clearly and concisely so that they can respond appropriately. Their proposal and your terms of reference will form the basis of the detailed plan for the specific evaluation.

**Your terms of reference should include:**

- *Why* and for *whom* the evaluation is being done – what will it be used for.

- *What* it intends to accomplish – what specifically you want to know.

- *How it* will be accomplished – the methods to be used. Note that you might not know this until you have appointed an evaluator. That's perfectly ok – it's what you will be paying them to do. In your RFP you should ask them to propose the methods and explain why they have selected them.

- *Who* will be involved in the evaluation, including the governance structure, how far stakeholders may be involved and how you will be supporting the evaluation. What you don't want is people to go away and not tell you anything until the end – you need to be involved in steering the work, without detracting from the expertise and independence of the evaluators.

- *When* milestones need to be reached including when the evaluation needs to be completed. Be clear on this, but be practical. Some activities take time – often more than you want or expect. If there are immovable deadlines, say so up front – a great report delivered too late is no use at all! Be responsive. If it transpires your data needs a lot more cleaning than expected, or it proves very difficult to track down past clients you may need to allow some extra time (and possibly budget).

Write a terms of reference document (or an evaluation brief) even if you are doing all the work internally. It's an important management document to make sure the evaluation actually delivers and doesn't get lost in the pressure of delivering your everyday work.

The Better Evaluation website has a tool to help you write a terms of reference. GeneraTOR is free software, but you will need to create an account. You can find more information here: betterevaluation.org/en/commissioners_guide/step3#anchor_generator

## Managing evaluations

Someone needs to manage the evaluation. They will oversee the development and implementation of the evaluation plan and help problem-solve if any issues come up. This person needs to ensure communication between the evaluators, the evaluation steering committee (if relevant) but also with the key stakeholders, such as the project/programme manager, and the relevant community. If your evaluation or the initiative is particularly complex, you may also have a technical advisory group of some sort. Some evaluations have a stakeholder group that can advise as the evaluation progresses. In some cases this is helpful to ensure buy-in.

One of the first tasks is to agree how frequently the evaluators should check in with the evaluation manager on progress. You also need a process for handling problems that may arise, especially if these will affect important delivery milestones, or where there may be communication issues with specific stakeholders. You may have specific protocols for interacting with stakeholders and clients – you will need to ensure these are followed and that the work is done ethically. If you are government funded and dealing with vulnerable populations, you may also need formal ethical approval for some of your data collection activities. Your evaluator or funding agency should be able to advise you on this. The ANZEA evaluation standards (see further reading, page 44) provide guidance on this aspect.

Ideally you will have agreed a fixed price (or investment if internal) for the evaluation. However, you may need to revise this if something significant crops up. This should be the exception, not the rule however. You will need to receive progress reports, notes on any implementation challenges and how these were resolved, and a revised evaluation work plan if that becomes necessary.

Managing an evaluation is an investment in getting the right product at the right time and on budget – and is often forgotten in the rush to appoint an evaluator.



*A good report will include clear and simple charts and diagrams that convey the key points clearly.*

## Assessing quality

The quality of an evaluation will depend on whether it has met the requirements that you set out (has it answered your questions?). It also depends on whether it was done well – was the methodology appropriate (scientifically and culturally), was it implemented correctly and are the conclusions well-founded.

**Things to look for in a good report will include:**

- An executive summary that contains direct and explicitly evaluative answers to the key evaluation questions used to guide the whole evaluation. This needs to get across the key points clearly and concisely as it will be the only part some people will read.

- Explicitly evaluative language (explains and judges) when presenting findings (rather than value-neutral language that merely describes findings) and provides examples.

- Clear and simple charts and diagrams that convey the key points clearly.

- Structuring of the findings section using the key evaluation questions as subheadings (rather than types and sources of evidence, as is frequently done). It should bring together several sources of information on each question as far as possible.

- Clear and transparent explanations of the reasoning used, that are understandable to both non-evaluators and readers without deep content expertise in the subject matter. These explanations should be broad and brief in the main body of the report, with more detail available in appendices.

- Just enough detail, with the technical details included as appendices. This should include the explanation of the methods used and the reason for using them.

## Using and communicating evaluation findings

Your evaluation report should be written in language that is appropriate and accessible to the key audiences of the report. This might mean it changes between the summary, the main body and the appendices. Or it may mean that several different reports/publications are provided for different audiences.

There are many ways to disseminate your findings – not just a report but workshops, seminars and conferences. It is important to share your findings. You will have wanted information on what other people are finding to inform your evaluation – they in return will want to learn from you. Sharing your findings publically also helps promote trust and support quality evaluation.

# Further reading and additional resources

**Evaluation Planning for Funding Applicants**
superu.govt.nz/resources/providers

**Evaluation Guide for Funders – How to work with providers to develop useful evaluation**
superu.govt.nz/resources/funders

**How to Develop a Monitoring and Evaluation Plan**
thehealthcompass.org/how-to-guides/how-develop-monitoring-and-evaluation-plan

**A Practical Guide for Engaging Stakeholders in Developing Evaluation Questions**
betterevaluation.org/en/resources/guide/engaging_stakeholders

**Writing a Terms of Reference for an Evaluation: A How to Guide** from the Independent Evaluation Group of the World Bank gives a detailed description of how to prepare an evaluation ToR and includes a checklist
betterevaluation.org/resources/guides/tor/how_to

**Evaluation Standards for Aotearoa**
superu.govt.nz/publication/evaluation-standards-aotearoa-new-zealand-evaluating-integrity

**The evaluation hikoi: A Māori overview of programme evaluation**
hauora.co.nz/assets/files/Māori/HMB_Māori-Evaluation-Manual-2009.pdf

**The Ultimate Guide to Effective Data Collection**
ebook.socialcops.com/data-collection

**A Step-by-Step Guide to Building a Monitoring and Evaluation Framework**
grosvenor.com.au/resources/building-a-monitoring-and-evaluation-framework

**Glossary**
A useful list of sometimes interchangeable terms and definitions
superu.govt.nz/resources/glossary

# Acknowledgements

This handbook is the result of input from too many people to name them all but we are grateful for all their comments and for allowing us to road test the handbook with them.
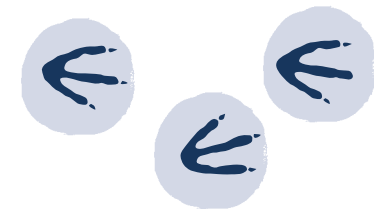
Finally, we thank you for your time and effort in exploring evaluation through this handbook. Your real evaluation journey starts now.

45

Although evaluation can seem daunting and resource-intensive, it doesn't need to be.

Start where you are, with small steps.

## About Superu

Superu is a government agency that focuses on what works to improve the lives of families, children and whānau.

**What we do:**

- We generate evidence that helps decision-makers understand complex social issues and what works to address them.
- We share evidence about what works with the people who make decisions on social services.
- We support decision-makers to use evidence to make better decisions to improve social outcomes.

**We also provide independent assurance by:**

- developing standards of evidence and good practice guidelines
- supporting the use of evidence and good evaluation by others in the social sector.

Social Policy Evaluation and Research Unit

superu

## Learn more at superu.govt.nz

Follow us on Twitter:
@nzfamilies

Like us on Facebook:
Social Policy Evaluation and Research Unit

**Families
Commission**
**Kōmihana ā
Whānau**