SIA

SOCIAL INVESTMENT AGENCY
ORANGA TANGATA

Investing in what works for better lives

An introduction to social investment analytics*

Measuring impact with integrated data

New Zealand Government

Creative Commons Licence

Liability

While all care and diligence has been used in processing, analysing and extracting data and information in this publication, the Social Investment Agency gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation

Intended audience

This paper is intended for analysts, policy teams, and managers who have an interest in social investment.

Resources, tools and guides

The Social Investment Agency is developing a range of tools, products and guidance to enable agencies to develop their social investment approaches, and analyse and measure the impact and effectiveness of the services they're delivering.

* By 'social investment analytics' we mean analysis of anonymised, integrated data with the objective of measuring the impact of government services on people's lives.

Turning data into insight

When we analyse integrated data it allows us to study people's journeys over time in order to learn what works, for whom, and at what cost. While this allows us to track people's journeys over time at the individual level, we do not report results for individuals, and we cannot see who these individuals are.

The analytic methodologies and 'big data' tools that support a social investment approach are powerful and innovative, but to be used effectively they require us to re-think the questions we ask of data. In the spirit of transparent, replicable, reusable and extendable work, the Social Investment Agency (SIA) is sharing what we've learned so that others can adopt these methodologies, if they are useful.

This document provides an introduction to the analytical methodologies used to inform decision making for a social investment approach. It outlines the data available, the insights we can generate and the types of questions we can seek to answer. This paper is intended for analysts, policy teams, and members of the public who have an interest in social investment and would like to know how social investment analysis can generate better insights into the population.

Integrated data gives us insight into people's lives

We integrate data about individuals so that we can better understand the complexity of their lives. When we use a single organisation's data to study the population they serve, we can't see much further than the outputs of the service provider and the trends indicated by annual snapshots and averages. A single-agency lens on a population prevents us from being able to see the compounding and interrelated factors that affect peoples' lives and needs.

We can also use integrated data to examine the effects of decisions in one area on outcomes in another, for example we may discover unexpected benefits or perverse outcomes. We can use this knowledge to make better decisions.

It's important to ask the right questions

In order to ask the right research question, we need to clearly define the problem we are trying to solve and the decision we need the analysis to inform. This guide covers the following questions that can be answered using social investment analytical techniques:

1. What are the characteristics of the population we want to better support?
2. How effective are specific investments for improving both short and long-term outcomes?

What data is available?

There are two main ways to access integrated data for analysis:

- Statistics NZ's Integrated Data Infrastructure (IDI).
- Matched data on individuals provided under Agreed Information Sharing Agreements (AISAs).

Statistics NZ's IDI

The IDI is a large research database curated by Statistics NZ containing information about people and households in New Zealand. This data comes from a range of government agencies, Statistics NZ surveys, and Non-Government Organisations (NGOs). All information in the IDI is anonymised - individuals cannot be identified. The IDI is a research database; it can't be used for operational purposes.

The data in the IDI can be used to understand individuals' interactions with services across the social sector, i.e. hospital discharges, enrolment and achievement in education, encounters with the justice system, etc. This detailed information allows analysts to piece together people's journeys over the course of their lifetime and analyse how their contact with services has or has not affected their path.

The IDI is an unstructured dataset, so the SIA have developed tools and resources to simplify analysis. Useful resources include:

- The Social Investment Measurement Map (SIMM), which provides information on what type of measures are available before accessing the IDI.
- The Social Investment Analytical Layer (SIAL), which simplifies the process of conducting an analytical investigation in the IDI by organising the unstructured data into usable tables.

Matched data

If organisations want to share their operational data, including identifiable data, they will need an Agreed Information Sharing Agreement, a contract, or a Memorandum of Understanding that sets out how the data will be collected, stored, analysed and used.

Ensuring the security and privacy of data is a key priority. Under the Privacy Act 1993 all organisations must have policies and procedures in place to provide the necessary assurance that data is only used or shared in a way that is permissible, and that organisations can guarantee the security of any personal information that is held.

What are the characteristics of the population we want to better support?

Descriptive analytics can provide us with useful information about a specific group of people.

We can learn about:

- Which characteristics (that can be seen in the data) are most strongly associated with an individual's likelihood to experience a given outcome.
- The existing service landscape and who is receiving which services.

We can seek to answer questions such as:

- What family and environmental factors are most closely associated with a young person disengaging from education?
- What early childhood indicators are associated with poor health later in life?
- What are the strongest risk indicators that a student won't achieve NCEA Level 2?

How this analysis works

There are many ways to approach analytically understanding a population; one is to begin with a fixed outcome that we want to understand the population's likelihood of experiencing, or not experiencing.

Understanding a population using data involves selecting an historical cohort – one for whom we have a sufficient timeline of longitudinal data for analysis – and calculating the relationships between the characteristics they exhibited at the beginning of the time period and the outcomes they were experiencing at its end. This can then be combined to identify the characteristics of the population group who are most vulnerable at an aggregate level.

In some circumstances, there may not be a fixed, measurable outcome that we want to understand. We may simply want to find out how known risk indicators are distributed throughout a population, and in what configurations they tend to cluster. In this case, we can choose to combine different sources of evidence by using indicators of risk identified by other analyses (such as the Growing Up in New Zealand longitudinal study), or similarly robust research projects, and test the population for the distribution and concentration of these variables. A further option is to then segment the population into sub-groups with similar characteristics and risks.

What data is required

Statistics NZ's IDI can be used for this work. The results will tell us 'who' at an aggregate level are likely to experience poor outcomes, but not identify them.

The following information needs to be clearly defined:

- The population of interest.
- The variables most closely associated in previous research with the likelihood of experiencing specific outcomes.
- The short-term outcomes of interest – if we are testing for a specific outcome.
- The long-term outcomes of interest – if we are testing for a specific outcome.
- The dates the analysis is required to cover.

What we can do with these results

Descriptive analysis helps us improve our understanding of the population group we are working with. Understanding the characteristics of the population can inform and complement an investigation into what causes the outcomes we are interested in, though it's important to remember that the characteristics identified in the data as being closely associated with an outcome are not necessarily the cause of that outcome.

Population analysis needs to be supported with strong contextual information, research and previously published literature before it can inform a service response.

Understanding the population through the lens of the services they receive can also be useful for learning which service touch points or trusted relationships might be the most effective conduits for improving support.

How effective are specific investments for improving both short and long-term outcomes?

Predictive analytics enable impact evaluations to empirically test whether or not an initiative has delivered the desired outcome, and for whom it works best.

We can learn about…

- Who is participating in current initiatives.
- Whether the desired population are actually receiving support and how effective this support is.
- The characteristics of the people for whom a specific initiative is most effective.
- The return in long-term value (fiscal, economic or social) of supporting an individual to experience a certain outcome.

All of the above can be combined to understand the value being returned.

We can seek to answer questions such as:

- Does a prison trades training programme improve rates of employment five years later compared with a control group that did not do the programme?
- Which groups most benefit from a given intervention?

How this analysis works

This analysis measures the effectiveness of initiatives. This is done after the initiative has taken place, by comparing those who received an initiative (the case group) with a control group who did not. The case and control groups are built in the data from the population of interest, controlling for every factor we can account for in the data.

A passage of time is required between the initiative being delivered and the outcome being sought so that measurement is of the same population who received the initiative. However, rather than waiting for outcomes to happen for a group who are currently receiving an initiative, we can measure outcomes over time by selecting a cohort who had previously received the initiative. We can do this by using matched longitudinal data to observe their outcomes after a period of time has elapsed. Results are compiled at the aggregate level, and individuals are not identifiable.

Comparisons can be made for both short and long-term outcomes, and the difference in outcomes between the case and control groups can then be attributed to the initiative. These results can be

combined with the results describing the general population to understand whether the population of interest are receiving initiatives or not. At an aggregate level, the effectiveness of these initiatives for the population of interest can also be understood.

In some cases it is possible to use cost data to establish the fiscal value of outcomes. If the value of outcomes can be calculated, it's then possible to calculate the long-term fiscal return on investment in an initiative.

What data is required?

The IDI can be used as long as participation data for the initiative of interest is available. Adding new information to the IDI can take Statistics NZ between several weeks and several months.

The following information needs to be clearly defined:

- The target population of interest.
- The variables that may identify increased likelihood of experiencing specific outcomes.
- The population's interactions with social services agencies.
- The short-term outcomes of interest.
- The long-term outcomes of interest.
- The initiatives to be evaluated.
- The cost of the initiatives.
- The dates the initiatives occurred.
- The dates the evaluation will apply to.

What we can do with these results

Using this kind of analysis, we can generate evidence upon which to base future decisions such as which initiatives to scale up or down, increase participation in, amend or performance manage.

The potential for using this sort of analysis to assess the effectiveness of initiatives should be considered when planning evaluations for new initiatives. Planning for evaluation before an initiative is implemented presents an additional opportunity to collect better data during delivery, and add this data to the IDI.

The SIA's Social Housing project is an example of this type of analysis in practice.