Technical Guide: Outcomes Measurement

For social impact investment proposals to NSW Government

July 2018

Office of Social Impact Investment

NSW GOVERNMENT

Contents

Glossary of key terms 4

1. Introduction 8

1.1 Context and purpose of this paper 8

1.2 What is in the guide? 8

1.3 What is not in the guide? 9

1.4 When to use the guide? 9

1.5 Relationship with evaluation 10

2. Program & measurement design 11

2.1 Introduction 11

2.2 Population 12

2.2.1 Identifying the target population 12

2.2.2 Expected effect of the intervention 14

2.2.3 Power and sample size 15

2.3 Intervention 16

2.3.1 Defining program logic 17

2.3.2 Key principles of program logic 18

2.4 Outcomes 19

2.4.1 Primary and secondary outcomes 19

2.4.2 Intermediate and proxy outcomes 19

2.4.3 Specifying outcome measures 21

2.4.4 Characteristics of outcome measures 21

2.4.5 Outcomes rate card approach 22

2.5 Counterfactual 23

2.5.1 Estimating the counterfactual 23

2.5.1.1 Counterfactual design options 23

2.6 Design elements 27

2.6.1 Statistical analysis 27

2.6.2 Acquiring data 27

2.6.3 Ethics 28

3. Valuing the outcomes – financial measurement & analysis 29

3.1 Introduction 29

3.1.1 Rationale for measuring outcomes 29

3.1.2 Financial measurement and analysis in context 29

3.1.3 Cost benefit analysis 30

Glossary of key terms

Term	Definition
Benefits	Benefits include financial, economic and social benefits of an intervention or program that can be used to support a business case for a social impact investment proposal. Benefits can be **direct** (e.g. immediate cash savings to the government) or **indirect** (e.g. avoided costs and productivity gains). **Intangible benefits** are those that cannot be measured directly in dollar terms (e.g. a community's increased trust in local police or a reduced fear of crime). In this guide, benefits refer to those that can be quantified and modeled in proposals. They are restricted to cash savings (current and future) and avoided costs that accrue to NSW Government agencies.
Cherry picking	Limiting services to recipients most likely to achieve positive change with the least intervention.
Cohort	A group of people to which another group of people is compared, according to some measure.
Counterfactual	An estimate of what would have happened in the absence of an intervention (a control group is often used).
Confounding	When a characteristic (called "confounder") is associated with both the intervention and the outcome of interest, and distorts the relationship between the intervention and the outcomes. Statistical techniques are available to adjust for known confounders during analysis. Randomly allocating individuals to the intervention and control groups is the only way to ensure that all potential confounders (i.e. those that are known and those that are not) are equally balanced between the two groups being compared.
Confidence grade	The extent to which benefits may be overstated and costs understated in a cost-benefit analysis. This may occur when data and evidence are uneven, old or incomplete. Over-optimism about the outcomes of an intervention, known as **optimism bias**, must be corrected in the analysis.
Confidence level	An estimate of uncertainty associated with the method used to create a sample of participants for an intervention. This sample should fairly represent the target population from which it is drawn. Statistical convention suggests that 90%, 95% or 99% confidence levels are acceptable levels of certainty. Setting confidence levels limits the likelihood of reporting false positives and/or false negatives.
Cost benefit analysis (CBA)	Analysis that comprehensively quantifies, in monetary terms, all the major costs and benefits of a proposal. Financial, economic and social benefits and costs should all be considered in CBA. They accrue to different people: some accrue directly to the user or provider of the service, while others will accrue to outsiders (these are known as "externalities").

Term	Definition
Costs	The financial, economic and social costs of an intervention. Types of costs include: • **Direct costs –** those that are directly related to a specific activity. General categories of direct costs include but are not limited to salaries and wages, fringe benefits, supplies, contractual services, travel and communication, equipment, and computer use. • **Indirect costs –** also known as overheads. Indirect costs refer to central administrative expenses, such as accounting and legal services, that are necessary for the continued functioning of an organization but cannot be directly allocated to a specific activity. They are typically allocated to a cost object on a systematic (transparent) basis. • **Intangible costs –** those that cannot be measured directly in dollar terms. Examples of intangible costs include pain and suffering, and lost confidence in the justice system. In this guide, costs refer to those to be modeled in proposals to form the basis of transactions, for instance: • set-up costs for the service (capital costs) • service delivery costs (e.g. staff salary and on-costs, overheads, etc.) • increased costs to other government services • transaction and evaluation costs.
Discounting	A method used to convert future costs or benefits to present values using a discount rate. **Discount rate** is the annual percentage rate at which the present value of a future dollar, or other unit of account, is assumed to fall away through time. In this guide, a central real discount rate of 7% is applied (potentially with sensitivity tests on the use of 4% and 10%).
Effect size	The percentage difference caused by an intervention, according to a reliable measure.
Historical Baseline	Historical data is analysed to establish the level of past outcomes for a cohort.
Impact	The longer term social, economic, and/or environmental outcomes (effects or consequences) of an intervention. Impacts may be positive, negative or neutral; intended or unintended.
Inputs	Resources put into an intervention for its establishment and implementation. Examples are money, staff, time, facilities, equipment, etc.
Indicators	Measurable markers that show whether progress is being made on a certain condition or circumstance. Different indicators are needed to determine how much progress has been made toward a particular goal, output, or outcome.
Intention-to-treat (ITT)	An analytic strategy to reduce selection bias, when comparing outcomes for an intervention group with those of a control group. All eligible people referred to an intervention are compared with eligible people not referred to the intervention, *regardless of whether those referred to the intervention actually receive it*. People who actually complete the intervention may differ in subtle ways (e.g. motivation to change) from those who are eligible for it but do not complete it or are not referred to it.
Intervention	A service or program of services designed to produce change in outcomes.

Term	Definition
Measurement *vs.* Evaluation	Measurement is the mechanism that tracks key indicators of progress over the course of an intervention as a basis on which to evaluate outcomes of the intervention. Metrics are the units of measurement. Evaluation is the systematic and objective assessment of the results of an intervention, particularly its effectiveness and efficiency. An evaluation framework details the method for collecting, analysing, and using information to answer questions about an intervention.
Monetisation	An approach to assign a monetary value to the social, economic and environmental costs and benefits in a CBA.[1]
Optimism bias	A demonstrated systematic tendency for appraisers to be over-optimistic about key project outcomes.
Outcomes	The changes that occur for individuals, groups, families, organisations, systems, or communities during or after an intervention. Changes can include attitudes, values, behaviours or conditions. Changes can be short term, intermediate or long term: • **Short term outcomes –** the most direct result of an intervention, typically not ends in themselves, but necessary steps toward desired ends (intermediate or long term outcomes). • **Intermediate outcomes –** link an intervention's short term outcomes to long term outcomes; they necessarily precede other outcomes. • **Long term outcomes** (sometimes called ultimate outcomes or impact) **–** result from achieving short term and intermediate outcomes, often beyond the timeframe of an intervention.
Outputs	The direct and measurable products of an intervention's activities and services, often expressed in terms of volume or units delivered.
Propensity Score Matching (PSM)	A statistical matching method used to estimate the counterfactual when random allocation to intervention and control groups in an intervention is not possible.
Proxy outcomes	A reliable indicator of an outcome that can be used in the absence of a direct measure when the actual measure is difficult to assess or occurs in the future.[2]
Perverse incentive	An incentive to act in a manner that goes against the desired objective of the intervention.
Power	The ability to find a statistically significant difference (i.e. a difference that is not likely to be due to chance) between groups (e.g. an intervention group and a control group) when one exists. Statistical power is a function of the effect size, the variability in the outcome, the confidence level, and the sample size.

[1] Monetisation is often used in frameworks based on Social Return on Investment (SROI). SROI is an approach to assign a monetary value to the social, economic and environmental outcomes created by an activity or an organisation. It is based on a set of principles that are applied within a framework (e.g. see The Social Audit Network manual available at http://www.socialauditnetwork.org.uk).

[2] Functionally, proxy and intermediate outcomes can be the same.

Term	Definition
Program logic	Presents the logic of how an intervention will work. The links between activities, intended outcomes, and between outcomes are shown, to articulate the intended causal links for the program. There is no one way to represent program logic – the test is whether it is a logical representation of the intervention's causal links. Synonyms include program theory, logic model, theory of change, causal model, outcomes hierarchy, results chain, and intervention logic.
Randomised design	Eligible individuals (or communities) are randomly allocated to either the intervention or the control group, whose progress is then tracked over time. Randomised designs have the advantage of avoiding selection bias in estimating the counterfactual.
Rate card	A list of outcomes government seeks to achieve and the price government is willing to pay for a successful outcome.
Risk	The likelihood that a particular event will occur. In this guide, risk is used to refer to likelihood of an adverse outcome of an intervention.
Selection bias	The systematic difference in characteristics between those who participate in an intervention and those who do not, thus affecting the validity of the comparison between the intervention and control groups. Bias may be due to (a) purposive program placement and/or (b) self-selection into the intervention. Bias can be due to observed characteristics, unobserved factors, or both.
Sample	A subset of the target population that provides a fair representation of the population from which it is drawn. "Fair" samples provide valid estimates of the population characteristics that they are supposed to represent if the findings and conclusions are to be extended to the target population at large (or 'generalised').
Target population	A group of people identified as having a set of shared characteristics and at whom an intervention could be aimed.

1. Introduction

1.1 Context and purpose of this paper

The NSW Government committed to develop guidance on measuring social outcomes as part of its *Social Impact Investment Policy*.

This guide aims to support proponents to develop a rigorous and appropriate measurement framework for, and demonstrate the value of, their social impact investment proposal. It has been prepared specifically for social impact investments with the **NSW Government** only. It is not intended to specify or guide measurement in other impact investments that do not involve the NSW Government and may not require the same approach.

As suggested by its name, this guide provides technical advice on designing an outcomes measurement framework, and on making the economic and financial case for an intervention financed by a social impact investment. It assumes readers have experience in and familiarity with statistical concepts, evaluation methods, and financial analysis. Where proponents do not have this expertise, you may wish to consider engaging external support to develop this part of your proposal.

This guide should be read with the *Principles for social impact investment proposals to the NSW Government*, which outlines five principles that every proposal should demonstrate. Two of these principles – **robust measurement** and **value for money** – are particularly relevant to this paper. The paper also touches on a third principle: **a service likely to achieve social outcomes**.

1.2 What is in the guide?

This technical guide outlines how to articulate the relationship between what an intervention is trying to achieve, how the intervention is going to achieve it, and how to measure the extent to which the intervention achieves its aims.

The guide is based on established frameworks and a large body of tested and accepted methods (in epidemiology, evaluation, and health and social economics), which have been tailored specifically for use in this context.

It is not an exhaustive collection of established approaches. Rather, the core chapters are designed to signpost key concepts that need to be part of the thinking underpinning development of social impact investment proposals. This includes:

- Developing a program logic that tells the story of the proposed intervention, how and why it will work, the goals of the intervention, and the process by which they can be achieved (Chapter 2).
- Designing the intervention and measurement framework in a way that will demonstrate the outcomes of the intervention compared to what would have happened in its absence (Chapter 2).
- Selecting appropriate outcome measures to demonstrate the impact reflected in program logic and intervention design (Chapter 2).
- Demonstrating value for money by attaching a monetary value to the benefits and costs of the intervention (Chapter 3).

1.3 What is not in the guide?

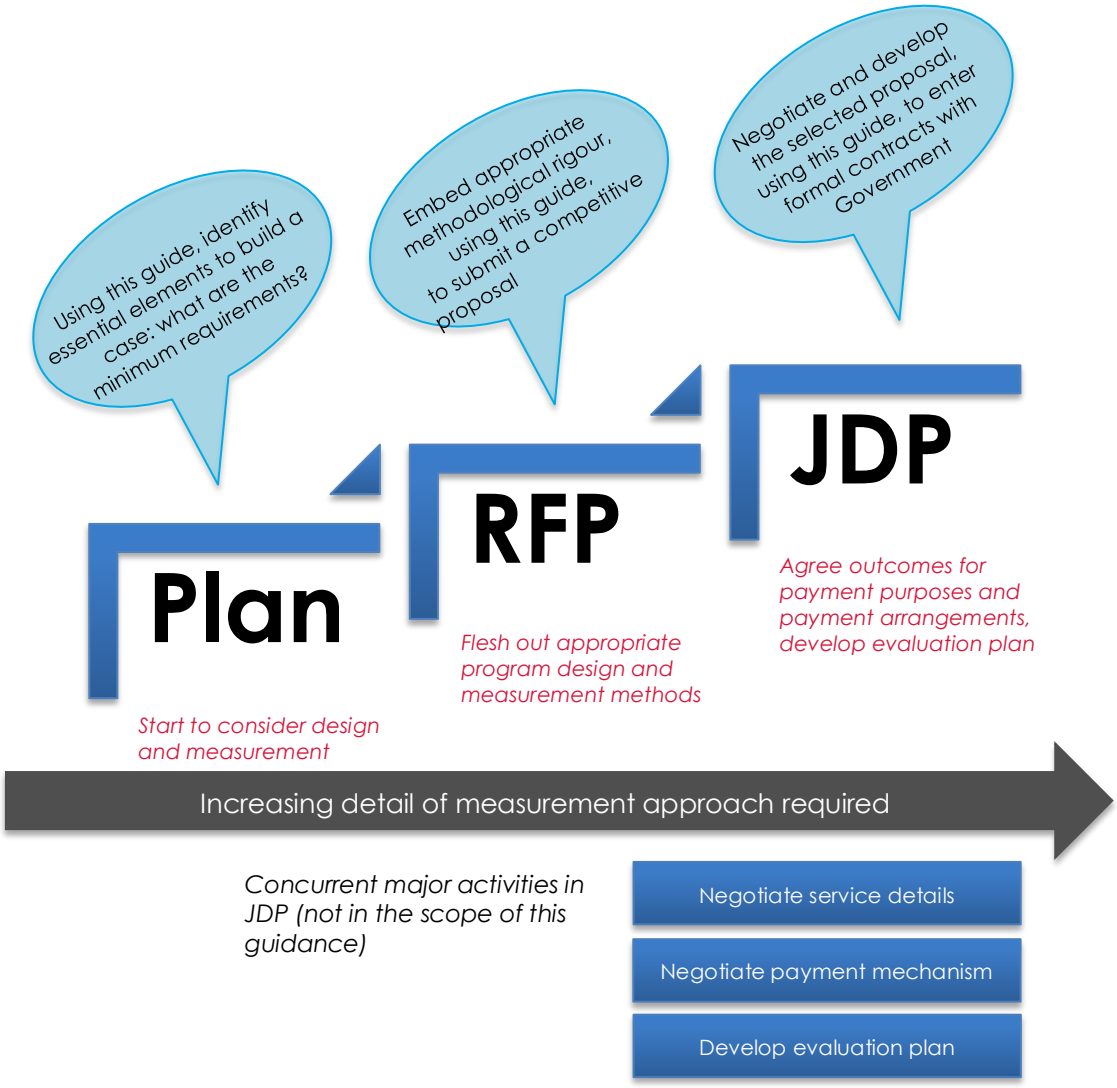
This guide is not a complete account of everything you need to develop a proposal. For example, it does not cover:

- design of programs or service models
- innovation in service models, financing and measurement design
- appropriate sharing of risk and return between parties to a social impact investment
- benchmark costs or key outcomes sought in policy areas
- evaluation plans.

1.4 When to use the guide?

This guide is useful at many stages of developing a proposal, from initial planning through to the joint development phase (JDP) (see Figure 1 below). We acknowledge that you may not be able to develop a measurement framework in your proposal to the standard set out in this guide. This may be due to poor data availability, constraints on the length of the proposal, or limited resources. However, you should aim to consider the issues canvassed here, and address them as best you can in your proposals, as they will most certainly need to be resolved in the JDP.

Figure 1: When to use this guide?



1.5 Relationship with evaluation

The chapters in this paper are essential precursors to designing an evaluation and the evaluation framework:

- Evaluation is defined as a rigorous, systematic and objective process to assess an intervention's effectiveness, efficiency, appropriateness and sustainability. Evaluation plays a key role in supporting decision making by helping understand whether an intervention is working, in what context, when it's not, and why. Well planned and executed evaluation provides evidence for improved design, delivery, and outcomes. The three main components of program evaluation are (1) process, (2) outcome, and (3) economic.
- Indicators need to be determined to effectively measure how much progress has been made toward a particular goal, output, or outcome.

While guidance provided here may inform an evaluation plan for an intervention, designing an evaluation is a completely separate process to designing an outcome measurement framework and should be carried out by independent evaluators. Further guidance on evaluation in NSW can be found in the NSW Government Evaluation Guidelines.

More information and guidance on evaluation is available in the NSW Government's Evaluation Toolkit.

2. Program & measurement design

KEY POINTS:

- A proposal must clearly identify the target population of the intervention and describe the criteria to define the intervention group.
- The overall logic of how the intervention is expected to work (i.e. the program logic) needs to be clear and based on quantitative evidence of its effectiveness.
- The primary outcome measure must be objective, reliable and collectable, and be linked to the social and financial benefits of the intervention.
- Outcome definitions should specify with what, when and how outcomes will be measured.
- The sample size should ideally provide at least 80% power to detect the effect, if any, of the intervention.
- A randomised design is the most robust way of assessing an intervention's impact.
- When randomisation is not possible, every effort should be made to create a control group that is as similar as possible to the intervention group and collect information on potential confounding factors. Where a control group is not possible or appropriate in the circumstances, other counterfactuals such as a historical baseline should be explored.
- Proposals should also discuss data management and ethics implications.

2.1 Introduction

This chapter provides practical guidance based on the first principle (robust measurement) in the *Principles for social impact investment proposals to the NSW Government.*

It sets out a framework for developing interventions proposed to be funded through social impact investment with the NSW Government. It provides guidance on how to design interventions so their effectiveness can be reliably measured and the associated social and financial benefits adequately quantified. This guidance assumes that proposals will put forward interventions with demonstrated efficacy (i.e. shown to have achieved outcomes under controlled conditions elsewhere), but will consider the scalability (wider rollout into business as usual) of the proposed interventions. This is important in building a compelling case for investment. However, we envisage that proposals are more likely to be a demonstration or proof of concept of working effectively at a small scale, at least initially.

The robustness and quality of measurement largely depends on the design of the intervention to allow effective evaluation, and relies on the four "*PICO*" pillars:

- **P**opulation
- **I**ntervention
- **C**ounterfactual
- **O**utcomes.

Proposals need to clearly identify the population targeted by the intervention, the details of the intervention being considered, the counterfactual, and the intervention's anticipated impact on

outcomes. Sections 2.2 to 2.5 address each of these four points, while Section 2.6 discusses issues related to data collection, analysis and ethics.

2.2 Population

2.2.1 Identifying the target population

Proposals should clearly identify the target population and how individuals will be selected to participate in the proposed intervention. Defining population characteristics forms the basis of eligibility criteria for the intervention. Among potential target populations, the levels of complexity, risk of adverse outcomes, and vulnerability may vary, and should be considered in the definition. Developing eligibility criteria for an intervention is important for two reasons:

1. To ensure a match between a particular intervention and those who are likely to benefit from it, on the basis of the stated social need. It is this group for whom interventions will be funded and outcomes improved. There must be clear criteria to identify the target population and a process to refer clients to the intervention.
2. To identify an appropriate comparison group. Both the intervention and comparison groups are drawn from the target population. Ideally, the members of the comparison group will have the same characteristics as the intervention group.

If the definition of the target population is not focused enough, the intervention may be too diffuse to have a significant impact on the target outcome. If the definition is too narrow, the target population may not be large enough to require a dedicated service or be generalised to a wider group.

Needs assessment is a systematic method to describe and characterise the target group against objective and detailed eligibility criteria. It generally includes descriptive historical data, for example:

- Trend analysis of proposed target population care flows (e.g. flows into care by age and referral type, care placements).
- Care journey analysis (e.g. historic trends on length of overall care journey, mix of placement types).
- Cost analysis (e.g. costs of typical care journeys, overall expenditure, and key cost drivers).
- Pathway analysis (e.g. referral pathways, current service user journey).

In most cases, proposals will seek to identify a subset or sample of the eligible population (i.e. the intervention group), rather than the entire population, to participate in the intervention. Particular attention should be given to the sample selection process. The process for selecting a sample should be as objective and systematic as possible to avoid introducing bias from self-selection or “cherry-picking”, and to ensure that the selected sample provides a fair representation of the population it is drawn from.

Box 1: Selection bias[3]

Selection bias describes the systematic difference in characteristics between those who participate in an intervention and those who do not, thus affecting the validity of the comparison. "Fair" samples must provide valid estimates of the population characteristics that they are supposed to represent. Only then can the findings and conclusions be extended to the target population at large (referred to in statistics as 'generalisability').

Randomly allocating participants to the intervention and control groups, with adequate concealment of allocation, protects against selection bias. Other means of selecting who receives the intervention, particularly leaving it up to the providers and recipients, are more prone to bias because decisions about eligibility can be related to perceptions about the intervention and responsiveness to it.

Examples of selection bias include:

1. Selecting volunteers into the intervention group and non-volunteers in the control group. Volunteers could be more change-ready than non-volunteers, resulting in greater impact of an intervention, such as improving parenting skills.
2. Studying the health of workers in a workplace compared to the health of the general population. Working individuals are likely to be healthier than the general population, which includes unemployed people (i.e. a healthy worker effect).

In addition, selecting a control group (see Section 2.5) should, as much as possible, mirror selecting those offered the intervention to prevent systematic differences between the two groups. The best way of achieving this is through a randomised experiment where the eligible population or sample is randomly split between the intervention and control group (see Factsheet 3). In practice, a randomised experiment may not always be appropriate or feasible and a range of other options could be considered (see Section 2.5).

Proposals should also discuss the processes anticipated to obtain consent to enrol participants in the intervention and/or acquire data. Section 2.6.3 discusses consent and ethical implications in more detail.

Each proposal should describe in detail the characteristics of the target population, including a list of eligibility criteria and the anticipated recruitment or referral process (i.e. how clients will be identified and engaged in the intervention). An example is provided in Box 2 below.

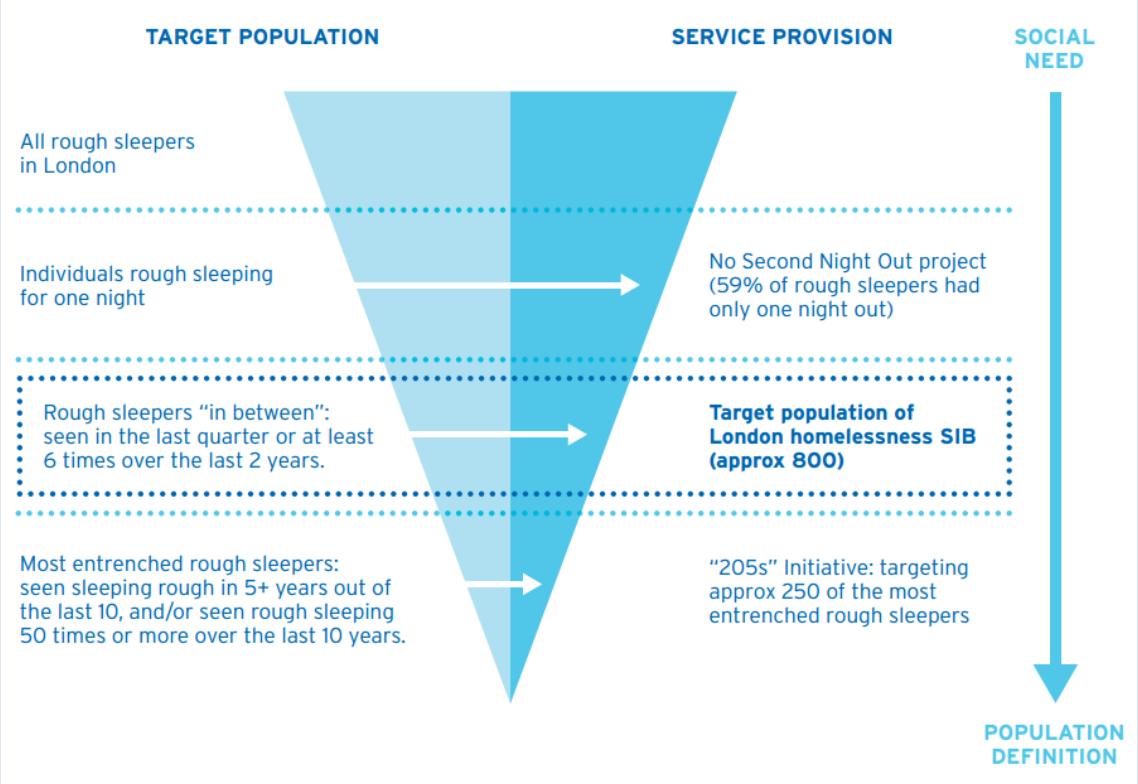
[3] Adapted from: http://www.medicine.ox.ac.uk/bandolier/booth/glossary/selectbi.html
http://ocw.jhsph.edu/courses/FundEpiII/PDFs/Lecture18.pdf

Box 2: Example of identifying the target population[4]

London's social impact bond (SIB) for rough sleeping ran for three years. It sought to improve outcomes for 831 people who move in and out of rough sleeping, and tackle the fundamental issues that often prevented them from benefiting from existing services.

The cohort comprised Londoners seen bedded down on the streets in the previous quarter, or living in a rough sleeping hostel and seen bedded down on the streets at least six times over the previous two years in the Combined Homeless and Information Network (CHAIN) database. CHAIN is a comprehensive database that records individuals' demographic information, support needs, and movement in and out of rough sleeping and hostel accommodations. The database is unique to London.

The diagram below illustrates an identified gap in services specifically targeting the "in between" rough sleepers.



2.2.2 Expected effect of the intervention

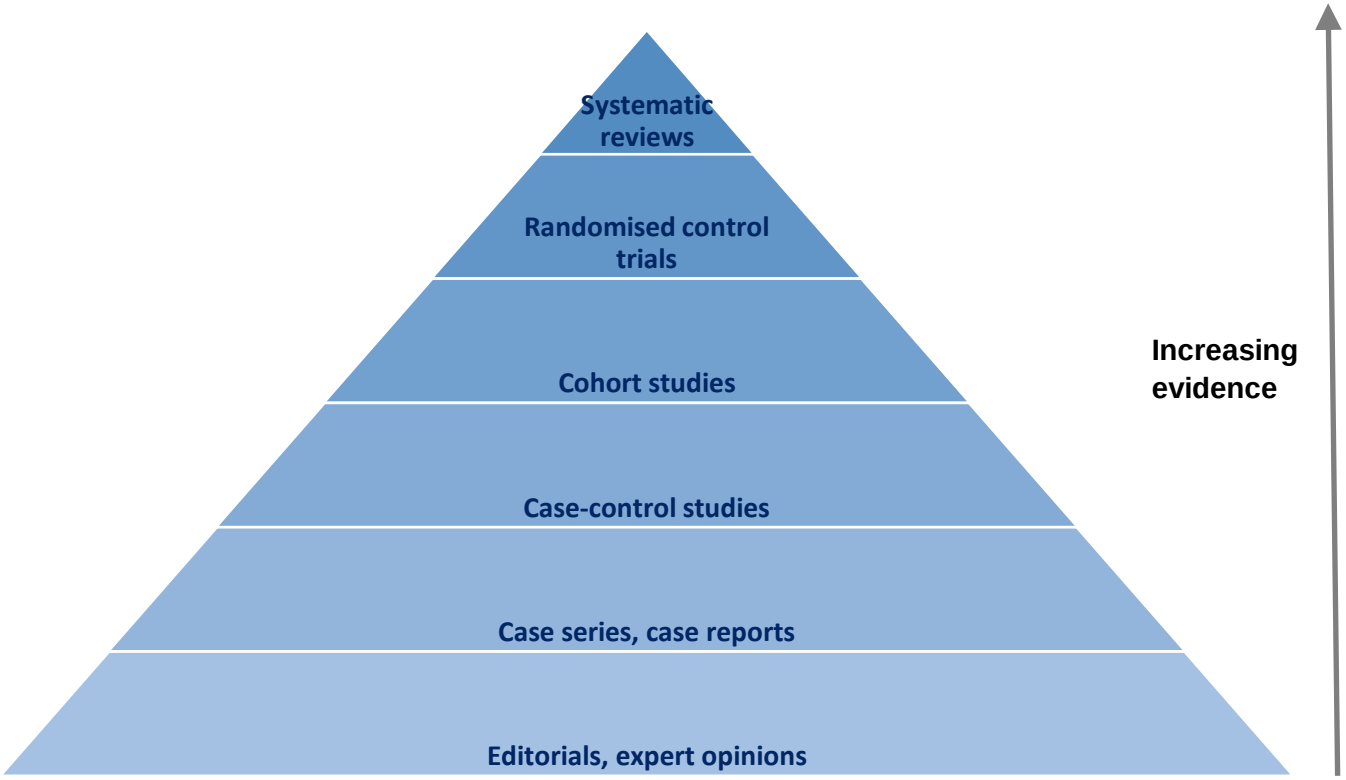
Proposals should establish what is already known about the intervention and state the size of the change in the outcome(s) the intervention is expected to have. That is, the difference one might expect between a group receiving the intervention and a similar group not receiving the intervention, also called the counterfactual (see Section 2.5).

The anticipated effect of the intervention should be based on a thorough review of the current evidence and reflect the degree of uncertainty associated with different sources of evidence.

[4] Ivy So and Adam Jagelewski (2013). *Social Impact Bond Technical Guide for Service Providers*. MaRS Centre for Impact investing. November 2013

Systematic reviews and large scale randomised trials provide the strongest evidence, while case reports and opinions provide the weakest evidence (see Figure 2 below). Effect size needs to be realistic and will ideally be based on prior research.

Figure 2: Hierarchy of scientific evidence



2.2.3 Power and sample size[5]

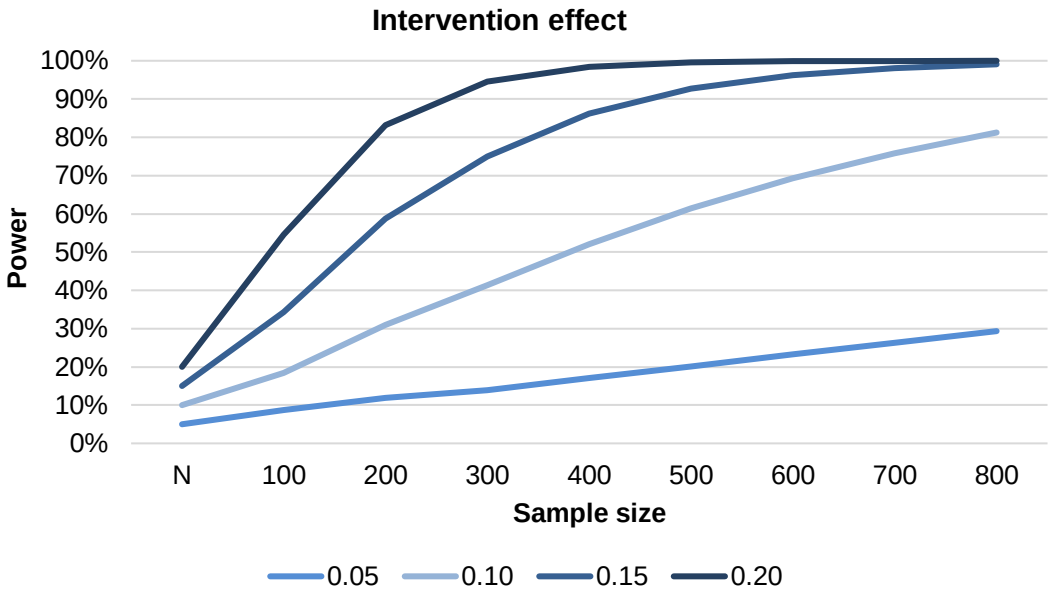
Power is the likelihood of detecting an intervention's effect when such an effect truly exists. It is affected by the size of the effect and the size of the sample used to detect the effect. Generally, larger effects are easier to detect than smaller effects, while large samples offer a higher chance of detecting an effect compared to small samples. As a result, both the effect size and sample size need to be considered when designing an intervention.

Power is expressed as a percentage, with larger values indicating a higher likelihood of obtaining statistically significant results. Larger values of power are desirable, with *at least* 80% being considered ideal for social impact investment proposals to the NSW Government. This means the proposed sample size for the intervention should have at least an 80% chance of correctly identifying an effect if it exists. While there are a number of online calculators available, various parameters need to be considered when calculating power.

The type-I error rate, commonly called "alpha", is the risk of reporting an effect when one does not exist (i.e. a false positive). The type-I error rate should be set at 5%. Figure 3 below illustrates the relationship between sample size (x-axis) and power (y-axis) for a binary outcome with the type-I error rate fixed at 5%. The power is based on a test comparing the proportion of individuals with the outcome in the intervention group to the proportion in the control group.

[5] Cohen, J (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. NY:Lawrence Erlbaum

Figure 3: Power and sample size



For the purpose of this example, we assume that the proportion of individuals experiencing the primary outcome in the absence of intervention is expected to be 50%. Each curve corresponds to a different intervention effect: the four scenarios assume the impact of the intervention increases the difference between the groups from 5% to 20%. Under all four scenarios, power increases as the sample size increases. However, for a given power, a much larger sample size is required to demonstrate a smaller intervention effect. In practical terms, this means that, with a sample size of approximately 200 participants, we would have an 80% chance of detecting a large impact of the intervention (say 20% difference due to the intervention) but only a negligible chance of picking up a small impact of the intervention (say 5% difference due to the intervention).

While a larger sample size increases the chance of demonstrating an effect, bigger is not always better. As seen on the curve corresponding to an effect of 20%, in this example, there would be very little increase in power in increasing the sample size beyond 400 individuals and this would increase service costs and resources. A very large sample size might provide enough power to detect very small differences. For example, 3,200 subjects would provide 80% power to detect a difference of 5%. However, unless a difference of 5% is deemed sufficiently important to influence future practice, exposing more participants than necessary to an intervention to show a difference that is too small to matter could be a waste of resources. Conversely, a design that does not have enough power to show a meaningful difference is likely to be inconclusive and also a waste of resources.

In addition, depending on the type of outcome and the intervention design, other parameters may need to be accounted for and identified in the calculation including the variance of the outcome, the duration of recruitment and follow-up, and the expected proportion of participants who might drop out of the intervention. Expert assistance may be needed.

2.3 Intervention

Program logic is used to tell the story of how an intervention works and why. When done well, it provides a clear and credible account of impact, setting out why the intervention is expected to have a positive effect on the outcome. It should explain why the impact of the intervention is

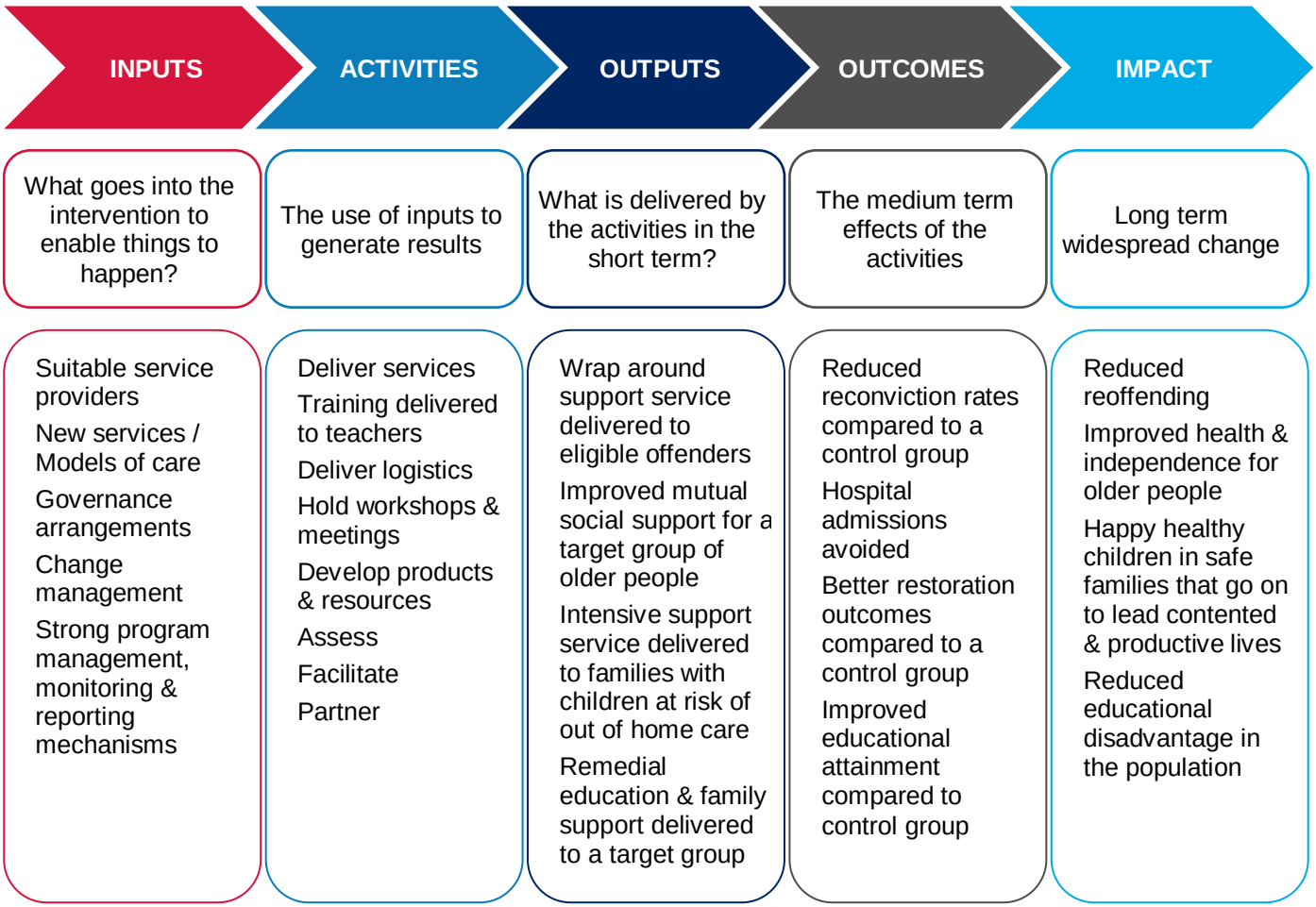
expected to go beyond what would have happened without it and why it is expected to improve outcomes compared to business as usual or competing interventions (if any).

By identifying the clients' needs and the effect the intervention is expected to have on those needs, program logic points to what should be measured to demonstrate that the expected impact is actually being achieved. These might be intermediate outcomes that lead to several others, or outcomes that make your intervention different from the usual practice. If measurement is not based on robust program logic, it risks not measuring the most important things and wasting resources.

2.3.1 Defining program logic

Program logic can be defined as a visual representation of how an intervention works. It describes the investment into the intervention, the strategies and activities to implement it, and expected achievements in the short, medium and long term. These components of an intervention are assembled into a causal chain that shows how the activities are assumed to contribute to immediate outcomes, to the intermediate outcomes, and to the longer term outcomes and desired impact (see Figure 4 below).

Figure 4: Program logic examples



2.3.2 Key principles of program logic

Program logic is a tool to bring rigour to crystallising the key aspects of an intervention and measuring its impact.[6] Different terms are used for this tool, including program theory, logic model, theory of change, causal model, outcomes hierarchy, results chain, and intervention logic.

However, the key principles to constructing program logic remain the same:

1. Define the purpose and objective of the intervention.
2. Bring together existing evidence about the proposed intervention: how and why is it expected to work?
3. Interrogate how the intervention in the proposed setting is expected to have an impact:
    - What is the path from the need you are trying to address to the change you want to achieve?
    - Are the goals / outcomes realistic?
    - Do the activities / interventions make sense, given goals / outcomes?
    - What are the assumed links between activities and outcomes?
    - How are outcomes connected?
    - How would progress towards the goals / outcomes be measurably demonstrated?
    - What are the hidden assumptions?
4. Based on identifying how the intervention has an impact, identify what should be measured to provide quantitative evidence of impact.
5. Identify measurable indicators that are sensitive to the activities of different actors and their outcomes.

There is no one way to represent program logic. Sometimes they are shown as a series of boxes (inputs → processes → outputs → outcomes → impact), sometimes they are shown in a table, and sometimes they are shown as a series of results with activities occurring alongside them rather than just at the start. The test is whether it represents the intervention's causal links and whether it communicates these adequately to the intended audience.

As an example, a theory of change may be developed to represent a program theory. It describes and illustrates how and why a desired change is expected to happen in a particular context. It is focused on mapping out or “filling in” what has been described as the “missing middle” between what an intervention does (its activities) and how these lead to desired goals being achieved. It does this by first identifying the desired long-term goals and then works back from these to identify all the conditions that must be in place (and how these relate to one another causally) for the goals to occur.[7]

For the purposes of social impact investments with the NSW Government, an approach based on describing a results chain (also known as a ‘pipeline model’) is particularly useful. It shows a program as a series of boxes [input → activities → outputs → outcomes → impacts] and depicts the outcomes leading up to the final impacts of an intervention. It can also include hypothesised causal links. Many interactive web-based tools are available to assist with this common approach.[8]

---

[6] Kellogg, W. (2004). *Logic model development guide*. Michigan: WK Kellogg Foundation.

[7] http://www.theoryofchange.org

[8] For example: https://fyi.uwex.edu/programdevelopment/logic-models/

Applying program logic to social impact investment proposals is illustrated by two current real-world exemplars.

📖 See **Fact Sheet 1** for measurement case studies.

2.4 Outcomes

Measuring impact is at the heart of social impact investment. Just as financial investments are often measured by their dollar return, social impact investments require a 'metric' for investors and the government to see social impact. Identifying the measurable outcomes of an intervention is a critical part of developing a proposal.[9]

Outcomes range from the ultimate outcome used to quantify the definitive impact of the intervention, to intermediate and process outcomes that quantify the fidelity of implementation. These outcomes and the way they are connected to the intervention – called an outcomes hierarchy – should be defined by program logic (see Section 2.3). An outcomes hierarchy shows all the outcomes, from short to long term, required to bring about the ultimate impact of an intervention. The ultimate impact is usually much longer term and aspirational, for example, the eradication of a social problem.

📖 The potential benefits brought by an intervention through its measurable outcomes are used to establish its benefit-cost profile (see Chapter 3).

📖 See **Fact Sheet 2** for examples of outcome measures used in social impact bonds and payment by results arrangements internationally.

2.4.1 Primary and secondary outcomes

Each proposal should aim to identify a single primary outcome, which represents the most important measure of impact. The primary outcome should ideally be the ultimate outcome directly targeted by the intervention rather than intermediate or process outcomes. The expected change in the primary outcome should be used to guide the sample size calculation (see Section 2.2.3). It is also likely to form the basis of payments in a transaction.

A proposal may also include a set of secondary outcomes. Secondary outcomes are often important measures of the effectiveness of the intervention that complement the primary outcome. In this context, '*primary'* and *'secondary'* are technical measurement terms, and should not be taken to underestimate the value of the full suite of outcomes in any program.

As an example, an intervention could propose to divert adolescents with behavioural problems from long term care. If successful, such an intervention might include a range of benefits to the participating adolescents, their families and their communities. The primary outcome, or the most direct benefit of such an intervention, would be reduced out-of-home care placements. Secondary benefits (depending on the intervention) may include improved family wellbeing and improved educational achievement for participating adolescents.

2.4.2 Intermediate and proxy outcomes

It may be difficult to observe the ultimate outcome because of a limited timeframe or because of measurement issues. In that case, the aim could be to measure proxy outcomes that are known to

[9] Muir, K., and Bennet, S. (2014). *The Compass: Your guide to social impact investment*. Centre for Social Impact Assessment: Sydney

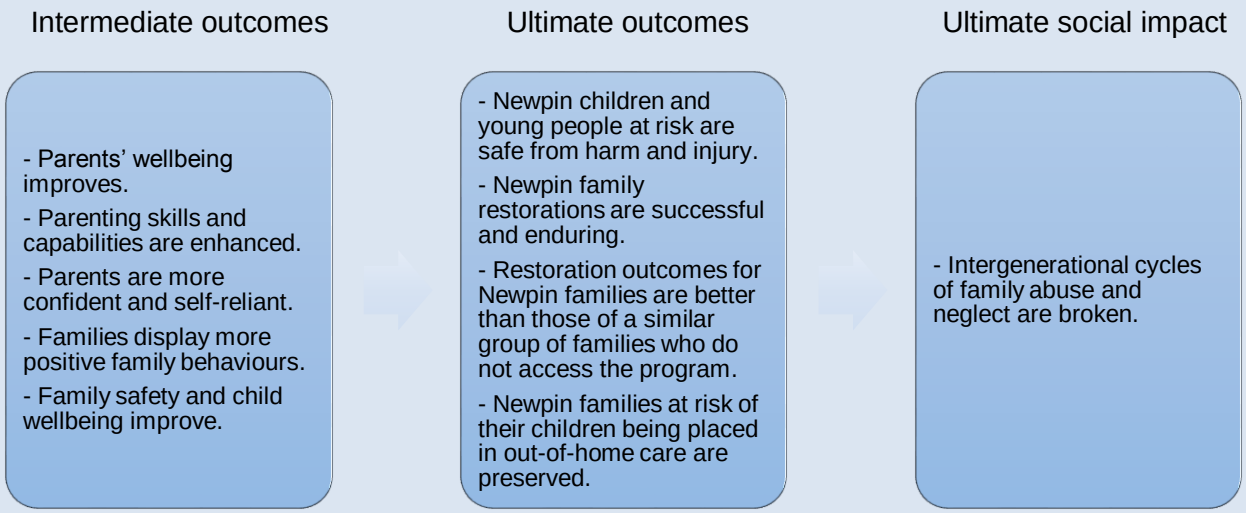
strongly predict the ultimate outcome. The strong predictive relationship of a legitimate proxy outcome and its measures with the ultimate outcome should be identified through review of the current evidence and take account of the degree of uncertainty associated with different sources of evidence (see Figure 2 above). This should be clearly identified in the program logic (see Section 2.3). For example, an intervention aims to reduce a certain type of cancer by introducing a new state-wide screening test. In this instance, it could take many years to observe a reduction in the number of cancers. Instead, assuming the proposed screening test has already been proven to predict reduced cancer cases, you could use the proportion of people undergoing the screening procedure as the proxy measure of the primary outcome for this program.

All programs will have intermediate outcomes. They represent progress along the outcomes hierarchy from short to long term. However, not all intermediate outcomes can be used as an acceptable proxy measure. To be acceptable as a proxy, the intermediate outcome must be a reliable indicator of the ultimate outcome. Reliable in this context refers to robust evidence (see Figure 2 above) showing that the proxy indicator predicts the ultimate outcome. Only then can it be used in the absence of a direct measure. Often, intermediate outcomes cannot sufficiently predict the ultimate outcome to replace it. An example might be rates of satisfaction with a smoking cessation program and rates of smoking cessation. While satisfaction with the program is an important prerequisite of cessation, it is neither the only way nor a guarantee of cessation. Although an important intermediate outcome, it does not sufficiently predict the ultimate desired outcome. In contrast, three months of abstinence from smoking might be a very strong predictor of cessation, making it an intermediate outcome that serves as an acceptable proxy for the ultimate outcome.

Achieving outcomes is the basis for making payments in a social impact investment with the NSW Government. We expect outcomes will be closely linked to the range of benefits an intervention aims to deliver. It can be difficult to measure social outcomes, particularly in the short term. However, proxy measures will need to be evaluated for correlation with the intended social outcomes as part of any proposal.

For example, Box 3 below describes the outcomes hierarchy for the Newpin social benefit bond. The bond aims to break intergenerational cycles of family abuse and neglect, and produce happy, healthy children in safe families that go on to lead contented and productive lives. This is impossible to measure over the seven-year life of the bond. Better parenting is an intermediate outcome, but does not sufficiently predict the ultimate social impact. On the other hand, whether children are in statutory out-of-home care or not is well-documented as a predictive intermediate indicator of ultimate social impact. In the context of social impact investment, improved restoration outcomes compared to those for similar families who do not have access to the program can be considered an acceptable proxy outcome of ultimate social impact.

Box 3: Outcomes for the Newpin social benefit bond



2.4.3 Specifying outcome measures

Outcomes can be measured in different ways. For example, in an intervention that aims to 'reduce traffic speeding offences', the outcome could be alternately measured as:

- the average number of new offences over a two-year period
- the proportion of individuals who commit a new offence
- the time taken to commit a new offence.

To reduce confusion, outcome definitions should specify the proposed measurement tool, the timing of the measurement, and the measurement method. For example, an outcome such as "improved health" is vague and requires more detail. "Improved health" could alternatively be defined as a reduced rate of hospital admissions over a two-year period following enrolment in an intervention. How outcomes are proposed to be measured will in turn impact on the statistical analysis of data (see Section 2.6.1) and sample size calculation (see Section 2.2.3).

In cases when more than one outcome is deemed appropriate, the most definitive one should be selected as the primary outcome and the other(s) as secondary. Definitive in this context refers to the outcome most accurately and specifically able to reflect the desired changes due to the intervention.

2.4.4 Characteristics of outcome measures[10]

Outcome measures are measurable markers that show whether progress is being made toward a particular outcome. Measuring outcomes should be based on indicators that have been shown to be reliable measures of effect and that are as objective as possible. Additionally, outcome measures must be available for participants from both the intervention and control groups, and be as complete as possible (i.e. minimal missing data).

The primary outcome must have an established reference value against, which it can be compared and costed for the benefit-cost analysis (see Chapter 3). For example, if the outcome is a reduced number of hospitalisations over a two-year follow up period, you would need to know the rate of

[10] See Fact Sheet 2 for desirable characteristics of outcome measures.

hospitalisation in the target population (e.g. among those older than 75 years living independently), as well as the costs associated with one hospitalisation for that group.

Both binary and graduated outcomes measures are possible, as long as they are robust and can be used to derive financial benefits. For example, a binary measure of recidivism may be whether a parolee reoffends within 12 months after release. Graduated measures could include a reduction in the seriousness of a re-offence, a reduction in the severity of a sentence, or a longer period from release to re-offence.[11]

2.4.5 Outcomes rate card approach

In some circumstance, government may provide greater specification of the outcomes sought. In an outcomes rate card approach, government identifies the outcomes it seeks to achieve and the price government is willing to pay for each outcome.

It is a tool that has been used by government overseas to develop multiple, outcomes-focused projects through a streamlined procurement process. The first rate card approach was developed by the UK Department of Work and Pensions in 2011, which stimulated 10 SIBs in 12 months. This approach has since been used to procure the majority of SIBs in the UK and has also been adopted in the US.

Box 4: NSW Homelessness Social Impact Investment Rate Card

OSII will pilot an outcomes 'rate card' as part of the 2018 Request for Proposals process to tackle homelessness. This responds to consistent market feedback on the need to streamline the transaction process, provide more data upfront and reduce the complexity of measurement frameworks.

The Homelessness Rate Card is presented below. It includes primary outcomes related to sustained accommodation, with secondary outcomes related to improved education/employment and reduced incarceration. The rates are provided as a guide to inform proponents' financial modelling, provide clarity around outcome metrics and provide a price signal.

Policy area	Metric	Cohort exiting custodial setting	Cohort exiting other government services
Housing	3 months sustained accommodation	6,500	4,000
	12 months sustained accommodation	10,000	6,000
	3 months sustained independent housing	16,000	11,000
	12 months sustained independent housing	16,500	12,000
Education / Employment	Training completion	3,500	2,500
	Engagement in structured activities	2,000	1,500
	3 months of sustained employment	4,500	3,500
	6 months of sustained employment	5,500	4,000
	12 months of sustained employment	6,500	4,500
	12 months of sustained employment	6,500	4,500
Justice	Not re-incarcerated 12 months after release	9,000	2,000
Maximum rate per client		$80,000	$51,000

Further information on the Homelessness Social Impact Investment Rate Card can be found in the appendix of the RFP documentation.

[11] Decisions about the measures most suited to deriving financial benefits in the context of Social Impact Investment transactions will be finalised during the JDP.

2.5 Counterfactual

One of the most important aspects in measuring the impact of an intervention is the ability to obtain a reliable estimate of the counterfactual (i.e. an estimate of what would have happened in the absence of the intervention). Proposals should consider how to assess whether any effects can be attributed to the intervention, that is, how you will measure how much of the outcome was caused by the intervention and how much was caused by other factors. The central feature of the counterfactual is that it constitutes an unambiguous and quantifiable estimate of the impact of the intervention.

For example, alongside a new cycling initiative there is a decrease in carbon emissions in a geographic catchment. However, at the same time, a congestion charge and an environmental awareness program begins in the catchment. While the cycling initiative may have contributed to reducing emissions due to motorists switching to cycling, the measurement approach needs to be able to determine the share of reduced emissions that can be attributed to the cycling initiative, rather than to the other initiatives.

2.5.1 Estimating the counterfactual

Estimating the counterfactual usually involves identifying an appropriate control group who did not receive the intervention. The similarity of the intervention and control groups is crucial. Ideally, the two groups will differ only in terms of whether they received the intervention or not so that any difference in outcomes can only be explained by the intervention itself.

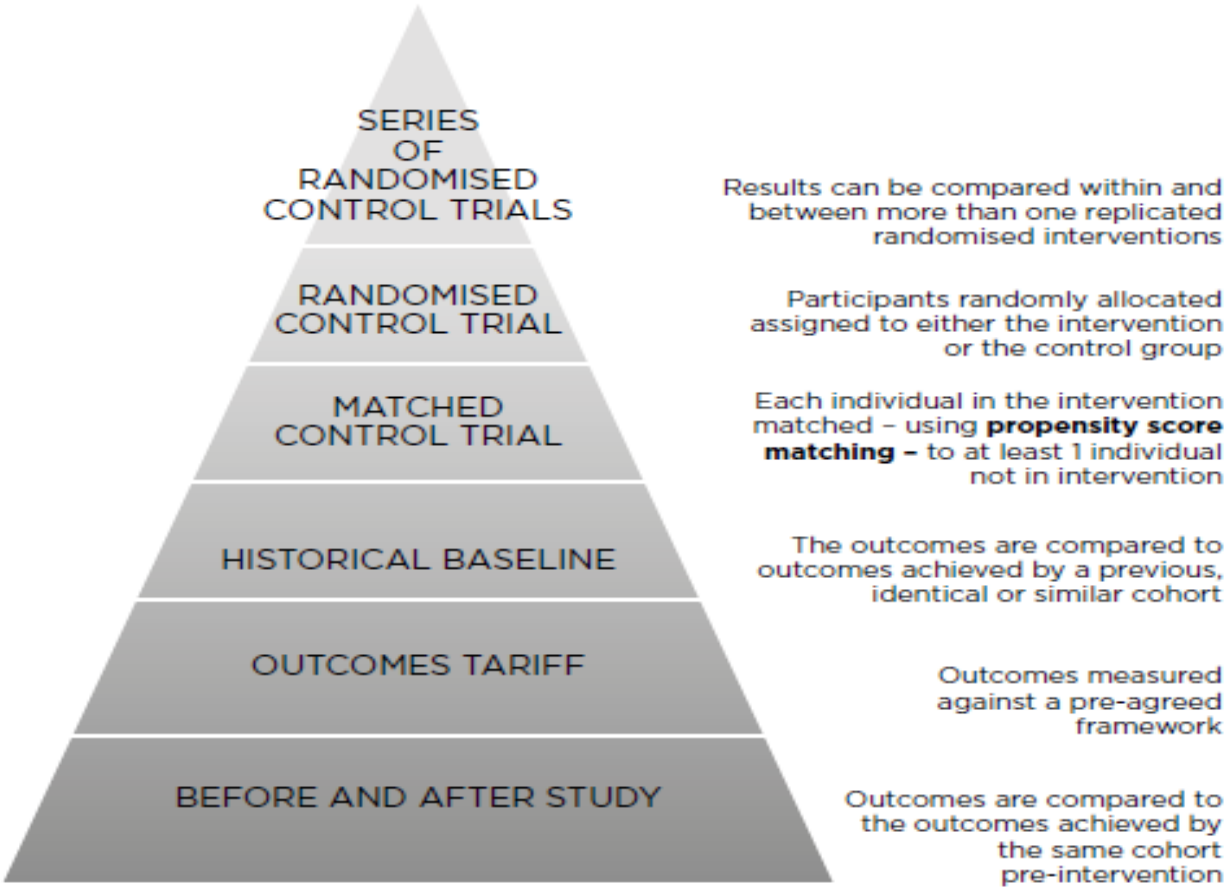
Similarity of the intervention and the control groups should not be taken to mean that intervention and control participants can be viewed as interchangeable. Those who actually complete an intervention may differ in subtle ways (e.g. motivation to change) from those who are eligible but do not complete it or are not referred to it. For this reason, cases cannot be reallocated: those eligible for an intervention who were referred but did not participate for some reason (e.g. refusal, compliance, dropping out) should not subsequently participate in the control condition.

2.5.1.1 Counterfactual design options

There are a number of designs available to estimate the counterfactual. Social Finance, UK has developed a hierarchy of these options, based on evidence quality of whether an effect is due to the intervention[12]:

[12] Social Finance (2015). *Technical Guide: Designing outcome metrics.* Social Finance: London

Figure 5 Hierarchy of Evidence Quality for Counterfactual Design Options

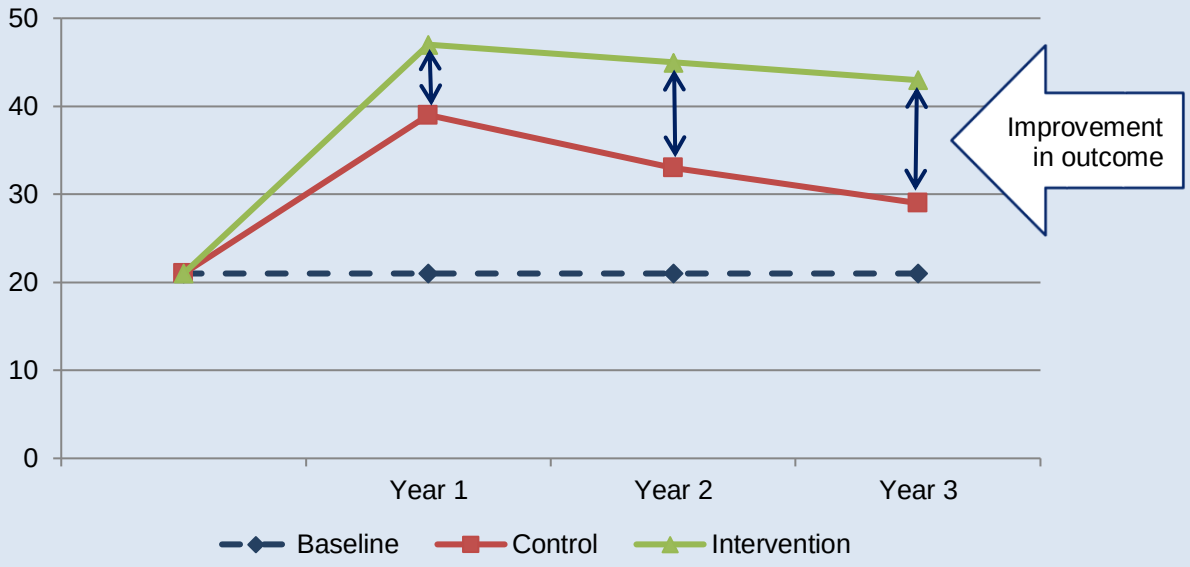


Box 5: Are all counterfactuals equal?[13]

In the graph below, the difference in an outcome between the intervention and control groups increases every year. Compared to the historical baseline, the intervention appears to produce a large change in the first year, but from the control group we can see that only a proportion of this change should be attributed to the intervention.

The top line (in green) illustrates some measure of the outcome of a successful intervention. The higher the outcome measure, the better it is. Compared to the control group, which did not receive the intervention, the effect of the intervention increased every year. However, if the intervention had been compared to a historical baseline only, its effect would have looked large in the first year, declining in the years after that. This example illustrates how changes to the environment may be captured by a control group, but not a historical baseline.

However, all counterfactuals will have advantages and disadvantages, which vary depending on the environment. A social impact investment may choose to accept this limitation of the historical baseline in an environment where there are no significant reforms occurring and the risks of using a historical baseline can be appropriately mitigated (e.g. by including timely baseline reviews).



Randomisation is the most robust way to determine whether an effect is due to an intervention. Ethical issues do not necessarily preclude the use of randomisation. Good randomised designs particularly for health and social care interventions take into account ethical considerations and have appropriate constraints built into the design.

[13] https://data.gov.uk/sib_knowledge_box/comparisons-and-counterfactual
Adapted from Haynes, L., Service, O., Goldacre, B. & Togerson, D. (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. London: Cabinet Office.

Box 6: NSW Social Impact Investments

To date, Social Impact Investments (SIIs) in NSW have tended to use randomised or matched control trials to measure outcomes. In order to simplify measurement, OSII is piloting a rate card approach in the homelessness request for proposals that is underpinned by a historical baseline.

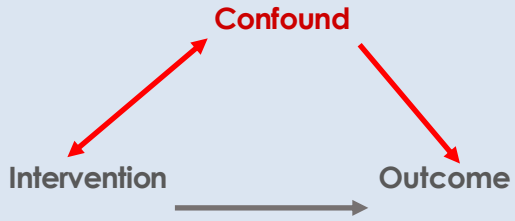
OSII acknowledges the lower degree of measurement rigour than randomised or matched control trials in this approach. However, this approach addresses market feedback for a SII development process that is less complex and accessible to more service providers, including smaller providers.

The rate card pilot will be complemented by a comparison group evaluation being conducted for the final evaluation. This evaluation will not only focus on the efficacy of the intervention(s), but also determine the effectiveness of a rate card pilot as a simplifying payment mechanism.

In practice, a randomised experiment may not always be appropriate or feasible. When randomisation is not possible, every effort should be made to create a control group that is as similar as possible to the intervention group and by collecting information on potential confounding factors (see Box 7 below). Determining the appropriate design requires consideration of cost, resources, quality of existing evidence and comfort levels of commissioners, service providers and investors.

Box 7: Confounding

Confounding occurs when a characteristic (called "confounder") is associated with both the intervention and the outcome of interest. This relationship is illustrated below.



This can occur when a characteristic present in the target population is not well balanced between the intervention group and the counterfactual. An example would be a health intervention where the most unwell individuals self-enrol to receive the intervention while others are used as the counterfactual. The most unwell end up being over represented in the intervention group than in the control group. Given the most unwell individuals are more likely to experience bad health outcomes, the effect of the intervention on the outcome becomes confounded by the health status of participants. In this case, it becomes difficult to separate the effect of the intervention itself from the confounding effect of the health status.

The best way to avoid this situation is to randomly allocate individuals between the intervention and control groups, thus ensuring that potential confounders are equally balanced between the two groups being compared and do not have the ability to distort the relationship between the intervention and the outcomes.

📖 See **Fact Sheet 3** on randomised and non-randomised designs, the advantages and disadvantages of each, and confounding factors.

2.6 Design elements

2.6.1 Statistical analysis

Proposals should briefly describe the intended method for analysing the outcome data. At a minimum, the main method for analysis of the primary outcome should be clearly specified. Details should include the population being analysed, the anticipated method(s) for dealing with confounds (e.g. randomisation, or propensity matching or historical baseline), and methods for dealing with missing data. Additional details may include possible subgroup analyses and potential sensitivity analyses.

For randomised designs, the main analysis should adhere to the intention-to-treat principle. This means analysing individuals according to the group they were allocated to, regardless of whether they ended up receiving the intervention (or control) as originally planned. In the case of non-randomised designs, a similar principle should be followed by not excluding any participant from analysis on the basis of noncompliance, protocol deviation, withdrawal, or anything happening after enrolment. This is because you would not know who in the control group might have dropped out had they the opportunity to receive an intervention.

2.6.2 Acquiring data

New data collection

When data required for measuring the impact of an intervention is not routinely collected or readily available, new data collection needs to be considered. You should carefully consider the exact type, format, and frequency of the planned data collection. Ideally, the costs associated with new data collection should be estimated and factored into the proposal. If not, the issue will be explored during the JDP. We encourage you to draw on existing datasets wherever possible.

Accessing existing data

In some cases, outcome data are already collected and can be accessed via data linkage. An example is health outcomes related to hospitalisations, which are routinely collected across NSW and held by NSW Health. In these circumstances, the process and costs for linking the data should be considered in the proposal, where possible. We envisage that information about potential data sources will continue to be refined during the JDP.[14] It may be necessary to obtain approval from appropriate data custodians to use existing data for the investment, which we will assist with.

In accessing existing data, proposals should also consider whether Minimum Data Sets[15] are available.

[14] It is recognised that assumptions will need to be made in the proposal regarding a range of variables, including data sources. For example, some agencies have linkage setup (e.g. BOCSAR), but others do not. Details of costs will depend on the ultimate design of the transaction agreed during the JDP.

[15] For example, http://www.adhc.nsw.gov.au/sp/minimum_data_set

Data management

Data collected for the evaluation of an intervention must be securely stored and adhere to individual privacy laws. For new data collection, it is important to think about the tools and processes that will be used to collect and store the data, and ensure the quality of the data.

Proposals should consider the need for data accessibility, use, and linkage. This is important when proposals anticipate using data from different sources, particularly outside of a single agency. As an example, linkage of any agency's administrative data to the Register of Births, Deaths and Marriages to determine mortality requires specific agreements with the Registrar at a Commonwealth level.

Proposals will involve a non-government third party, as a partner in the consortium and/or as an evaluator. There may be specific considerations that relate to third party access and use of data. Data quality

Proposals should also consider the quality of data that has been or is to be collected. A useful guide is the ABS Data Quality Framework.[16] The seven dimensions of quality are Institutional Environment, Relevance, Timeliness, Accuracy, Coherence, Interpretability and Accessibility. All seven dimensions should be included for the purpose of quality assessment and reporting.

2.6.3 Ethics

Most research proposals involving human participants need ethics approval and social impact investments may fall into this category. Proposals should clearly outline the ethical implications related to the intervention especially regarding the following:

- potential risks associated with the intervention
- process to protect individual data privacy
- consent processes to enrol individuals in the intervention
- consent processes to collect new data and/or access existing data
- methods to reduce perverse and unintended outcomes.

Proposals spanning more than one cluster (e.g. both education and health) may include more complex ethics considerations for data sharing and the like. Proposals should recognise the difficulty of data linkage for primary outcomes measurement due to the complexity of acquiring data across NSW Government clusters.

Note, individual agencies may have specific requirements to satisfy before granting access (e.g. the Department of Education requires researchers to complete a State Education Research Applications Process (SERAP) if research involves school-based activity). You will be required to comply with all agency-specific requirements.

Processes to obtain ethics approval or approval to access data held by government agencies are most likely to commence during JDP, though the likelihood of needing these approvals should be identified in proposals.

[16] http://www.abs.gov.au/ausstats/abs@.nsf/mf/1520.0

3. Valuing the outcomes – financial measurement & analysis

KEY POINTS:

- Financial cost benefit analysis is the valuation method preferred by the NSW Government to value the outcomes of social impact investment proposals.
- The costs included in the financial analysis are those involved in implementing the intervention, increased costs of other government services as a result of the intervention, and costs of administering the transaction and collecting data.
- The benefits include those that are cashable – that is, immediate savings to the NSW Government, in terms of reduced service demand and potential revenues from the intervention. Other types of benefits (e.g. long term cash savings and avoided costs to NSW Government) may also be considered.
- When a transaction involves investors (e.g. a social benefit bond), the total benefits must exceed costs to a degree that enables payment of investors' returns.

3.1 Introduction

3.1.1 Rationale for measuring outcomes

This chapter sets out a framework for measuring financial outcomes and impacts associated with interventions funded through a social impact investment with the NSW Government. It provides practical guidance and outlines considerations based on the second principle (value for money) in *Principles for social impact investment proposals to the NSW Government*. While the focus of much of the discussion is on all forms of social impact investment transactions, some issues highlighted relate specifically to social benefit bonds.

This chapter follows directly from Chapter 2, which described methods to demonstrate the effectiveness of interventions. As a general rule, a financial return is possible only if an intervention demonstrates it is effective. This means that interventions that cannot demonstrate a robust and statistically significant effect do not warrant financial valuation of the outcome. The further implication is that social impact investment proposals need to provide credible projections of the effectiveness of their intervention(s) to be able to forecast plausible financial outcomes.

3.1.2 Financial measurement and analysis in context

Financial measurement should essentially compare the state of the world with the intervention in place versus a status quo option – the state of the world without it. The nature of this comparison needs to be defined. This means specifying the intervention – its boundaries, the activities it entails, and the resources it consumes – and likewise, the comparison. This is important, as highlighted later, in estimating the costs of an intervention. To a large extent, the nature of the comparison will be determined by the counterfactual (see 2.5) used to demonstrate the intervention's effectiveness.

3.1.3 Cost benefit analysis

The defining characteristic of cost benefit analysis is that it values the costs and benefits of interventions in commensurate monetary terms.

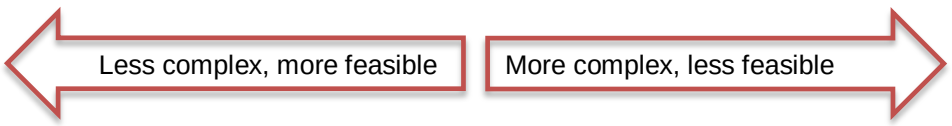
As costs and benefits are valued in the same (monetary) units, the advantage of cost benefit analysis is that it generally provides a clear decision rule for decision makers: if benefits exceed the costs, then the intervention with the highest net present value should be considered. In the context of social impact investment, valuing 'benefits' in monetary terms enables returns to investors and government savings to be clearly determined.

The main disadvantage is that monetising benefits can be difficult, particularly when the value of some outcomes is intangible (e.g. community or user satisfaction with a service).

Figure 6, below, illustrates different benefits and how they contribute to the complexity of social impact investments.

Figure 6: Different benefits and how they contribute to the complexity of social impact investment transactions

Nature of benefits	cash savings	avoided costs	productivity gains	other
Beneficiaries	a single government agency	multiple government agencies	other governments	individuals and communities
Timing of benefits	immediately	months	years	decades



3.1.4 Financial cost benefit analysis – a restricted version

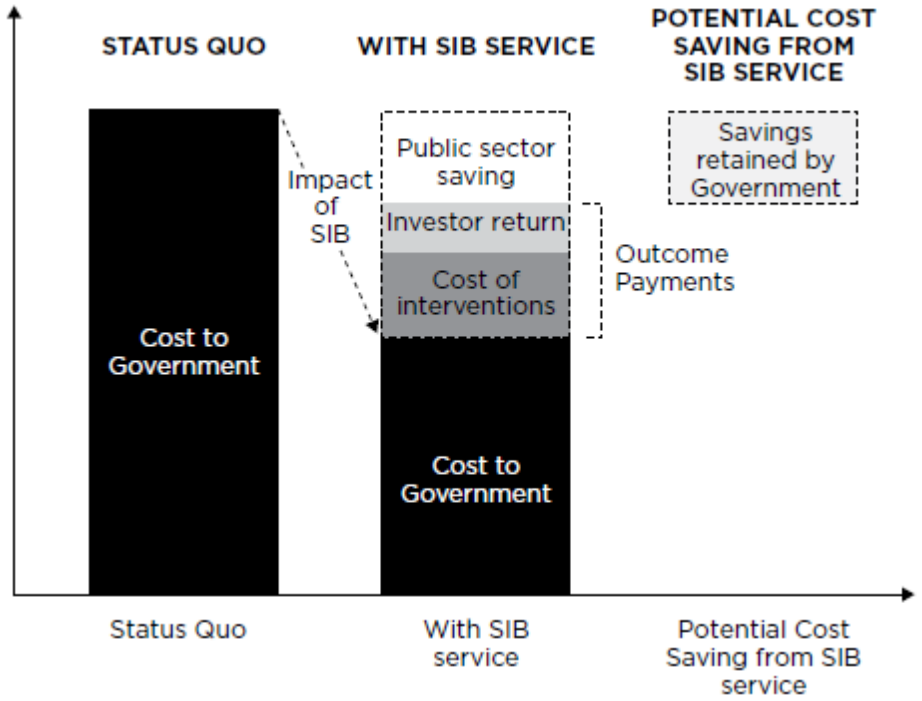
Given the difficulty for proponents to do a full cost benefit analysis, we consider a restricted version of cost benefit analysis that focuses on the financial position of the NSW Government (termed "financial cost benefit analysis" for the purpose of this guide) is sufficient for the purpose of demonstrating a proposal's value for money. Value for money is the basis of recommending NSW Government participation in social impact investments.[17]

There is existing NSW Treasury guidance on cost benefit analysis and financial appraisal. A full cost benefit analysis or other measurement methods may be complementary to the requisite financial analysis (see section 4.6).

[17] HM Treasury (2013), *Green Book Supplementary Guidance on Public Sector Business Cases Using the Five Case Model*, Lowe, HM Treasury
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469317/green_book_guidance_public_sector_business_cases_2015_update.pdf

Figure 7 below illustrates a simplified financial model of a social impact bond (known as a "social benefit bond" in NSW), based on the financial cost benefit analysis approach.

Figure 7: Visualisation of social impact bond financial model[18]



A robust financial model should account for:

- current and future costs to the government of a particular target population
- costs of a proposed intervention and the business as usual scenario
- estimated impact of the proposed intervention on current and future costs
- potential cost savings to the government as a result of the intervention.

Key issues in developing a financial cost benefit analysis include:

- Have you considered all the costs involved?
- Which government department(s) will bear any flow-on costs?
- Where do we get data to assess these costs?

3.2 Types of costs

In this context, costs included in the financial model should include:

- set-up costs for the service (capital costs)
- service delivery costs (e.g. staff salary and on-costs, overheads, etc.)
- increased costs to other government services
- transaction and evaluation costs
- any other costs not included above.

[18] Social Finance (2013). *A Guide to Social Impact Bond Development*. Social Finance: London

3.2.1 Recurrent versus capital costs

The collection of cost data is a task ideally planned at the outset of an intervention's rollout. It is important when costing an intervention to consider all the types of costs that may be incurred that are relevant to the intervention. These are costs incurred in implementing and running the intervention and fall under two broad categories:

Capital costs

These are initial costs of setting up an intervention and include items such as the purchase of equipment. Although the outlays for these items are often one-off at the outset of the project, their use may extend over a number of years. Such costs are fixed – which means that they do not increase according to use.

Recurrent costs

Recurrent costs are those costs that are required to operate an intervention. These include staffing and consumable items, such as medication, stationery and fuel.

Program records will be the primary source of data on these costs (e.g. staff salaries and expenditure on consumables). There may also be resources used and costs incurred from existing services in the running of the intervention. Consequently, these costs may not appear in the intervention's budget. For instance, a school based health promotion program might involve teaching staff delivering healthy eating and lifestyle messages. Although these salaries may not be directly paid from the intervention's budget, they nonetheless represent direct costs that need to be considered in the analysis. In this example, salary costs should be apportioned based on the time spent by teachers in delivering the intervention.

3.2.2 Costs of other government services

It is important that costs related to an intervention's referrals to other services are included in the financial model. For example, referral to other services, such as mental health or drug and alcohol treatment, may be needed to achieve intervention outcomes. These would generally be onward referrals from a case worker.

Where effective referral is vital to the success of an intervention, it is important that referral costs are accurately identified, analysed and attributed to the relevant supporting agencies. If this is not possible, a best estimate should be provided.

3.2.3 Transaction and evaluation costs

Transaction and evaluation costs need to be factored into the final financial model for the transaction (see Table 1). Proponents should make allowance for these costs, noting all are subject to refinement during a JDP.

These costs may include administration, contracting, professional services (e.g. legal and consulting advice), financial intermediaries (if applicable), data collection, and independent evaluation concurrent to service delivery. These costs are typically included in program evaluations and therefore the guidelines here are not reflective of standard practice.

Table 1: Summary of the different categories of costs

Type of cost	Description	Example
Capital costs	Upfront cost for an asset that has use over a number of years.	Equipment Land and building
Recurrent costs (fixed vs. variable)	Costs that are ongoing for the operation of an intervention.	Staff, consumables (such as stationery, medications, fuel).
Costs of other government services	The costs that flow on to other government services as a direct consequence of an intervention.	When implementing a school based health education program, one such cost may be the time spent by teaching in arranging for this program to be delivered.
Transaction and evaluation costs	These are costs associated with setting up the transaction and in its evaluation. Re-calibration of programs and services connected with implementing new practices; establishing partnerships where none existed before.	Costs of data collection and costs of engaging consultants to conduct the cost-benefit analyses or to act as financial intermediaries.

3.3 Benefit considerations for a financial cost benefit analysis

Key issues to consider:

- How will cost savings be achieved?
- How much will the government save if the outcome is achieved?
- Which NSW agencies will financially benefit if the outcome is achieved?
- What is the nature of the benefits?
- Where applicable, do the savings allow sufficient returns for investors?

3.3.1 Overview

Compared to costs, measuring and modelling benefits is generally more complex. Given this complexity, it is important for proponents to ensure benefits are not double counted in the proposal.

In the context of a financial cost benefit analysis, ‘benefit’ refers to the measurable benefits to the NSW Government in terms of:

- direct (or ‘cashable’) benefits, including revenues from the intervention and immediate savings
- long term cashable savings
- long term avoided costs
- productivity gains
- measurable social benefits and other benefits (i.e. the bottom two rows of Table 2, below).

Table 2 below summarises different types of benefits and how they contribute to the complexity of social impact investment proposals.

Among the most straightforward ways of funding financial returns and other costs is immediate cash savings to the government. We acknowledge that benefits may be dispersed across different government agencies. For example, an effective service to a homeless person may lead to savings in the housing, health, and police departments. Benefits may also accrue across different levels of

government (e.g. Commonwealth, local councils), such as when ex-offenders complete training that returns them to work (rather than reoffending). In this case, the NSW Government saves on prison costs and the Commonwealth Government saves on paying unemployment benefits.

In general, however, the wider the benefits are dispersed, the harder it will be to complete a social impact investment, with negotiation and partnerships required across different organisations, government agencies and jurisdictions. The benefits described in the yellow shaded boxes are those that should be considered for the purpose of financial measurement and modelling. A broader range of benefits may be explored during the JDP.

📖 Refer to *Principles for social impact investment proposals to the NSW Government* for a full discussion on the nature, recipients and timing of benefits of social impact investments.

Table 2: Different benefits and how they contribute to the complexity of social impact investment transaction

Increasing complexity of transaction

WHO RECEIVES THE BENEFIT?	Cash savings (future)	Avoided costs	Productivity / capacity enhancements	Other measurable benefits	Other measurable benefits
Community	• Effective crime prevention and re-offender strategies could reduce the need for businesses to pay for legal costs due to criminal activities.	• Better health outcomes could result in future savings for non-government organisations that provide non-health services (i.e. housing).		• Improved health and education outcomes could lead to better productivity and jobs, and more people able to participate in and their communities.	• Safer, more productive communities and better functioning families due to reduced antisocial behaviour. • Improved access to services for disadvantaged families and communities. • Greater transparency for taxpayers due to increased focus on outcomes.
Private individuals	• Effective crime prevention and re-offender strategies could reduce the need for private individuals to pay for damage to property and other costs (i.e. temporary vehicles due to car theft).	• Better literacy and numeracy outcomes could reduce the need for parents to hire private tutors for their child(ren) in the future.		• Improved health outcomes could lead to increased individual productivity. • Reduction in crime could reduce the level of lost productivity associated with the victims of crime (i.e. time spent in hospitals, fixing damage, away from work, etc.).	• Improved family functioning, relationships, health and wellbeing, employment opportunities, and living conditions. • Improved school attendance from better literacy and numeracy outcomes leading to better qualifications. • Better housing outcomes lead to better quality of life.
Other government – Commonwealth	• Positive change in outcomes for those accessing homelessness services could lead to reduced need for benefits (i.e. welfare).	• Better education levels, increased employment and reduced income inequality could lead to future savings in welfare payments.	• Decreased need for the Commonwealth to contribute to facilities for acute services.	• Improved health outcomes could increase individual productivity and reduce Commonwealth expenditure on welfare payments and intensive employment services	• Increased employment due to improved education outcomes could boost tax revenue.
Multiple NSW government agencies	• Lower recidivism rates could lead to cost savings for corrections, health services, police and court services.	• Increases in permanent supportive housing could lead to future savings for health, corrections and housing. • Reduced youth homelessness could lead to future savings from reduced hospitalisations and contact with the adult justice system.	• Improved education outcomes could reduce the demand not only for remedial teachers but also for new social housing units.	• Improved mental health outcomes could slow increasing demand for programs providing non-income support, disability and community services, housing and homelessness services, special schools and support classes, police, courts, prisons and juvenile justice. • Lower re-offending rates could help reduce cost pressures across criminal justice system including police, courts, legal aid, correctional services, juvenile justice and public prosecutions.	• Increased evidence base and availability of robust data for future policy makers as a result of the need for robust measurement. • Improved accountability for the effectiveness of expenditure on social services. • Limiting the risk to the government of funding ineffective programs.
Single NSW government agency	• Reduced care placements could lead to care cost savings. • Reduced homelessness could lead to savings in temporary accommodation costs.	• Reduced offending behaviour among adolescents could reduce local youth offending costs. • Savings to the government from reducing the number of children in out-of-home-care (through prevention and restoration) as they are not in long-term care. • Better literacy and numeracy could reduce the future need for remedial teachers.	• Improved health outcomes could lead to avoided capital costs for hospital and community care facilities (i.e. new hospitals). • Reduced offending could lead to reduced need for justice facilities.	• Reduced offending could lead to a more efficient police force. • Reduced homelessness could reduce pressure on outreach services. • Improved health outcomes could lead to more efficient hospitals. • Improved health due to increased physical activity levels could reduce pressure on treatment for diseases linked with lack of exercise.	• Accessing private capital facilitates upfront expenditure over and above what is available from public funds when expenditure is needed. • Better outcomes by providing a direct financial incentive for a service provider to focus on and improve the outcome in question. • Better evidence base for agencies on which services can achieve outcomes.
	TYPE OF BENEFIT				

Increasing complexity of transaction

3.4 Bringing it all together

3.4.1 Input data

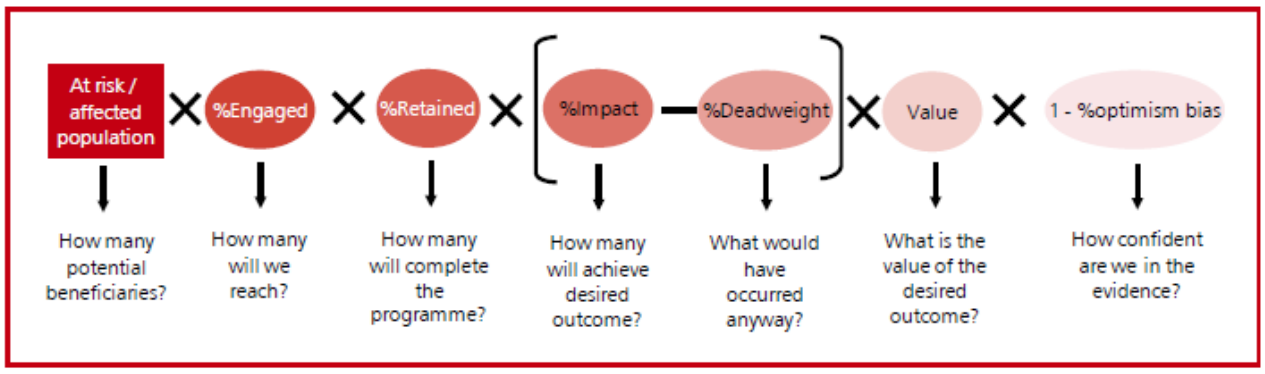
For each outcome, intervention data is needed to determine inputs for the financial cost benefit analysis.[19] Input data required include:

- total population at risk / affected (as defined by the eligibly criteria – discussed in Chapter 2)
- level of engagement with the target population (percentage of individuals who engage with the intervention)
- retention of the cohort (percentage of individuals who continue to be engaged until the intervention is complete)
- scale of impact in changing the outcome (success rate of achieving the desired outcomes – derived from measuring outcomes for the intervention group)
- what would have happened under business as usual (derived from measuring outcomes for the counterfactual as discussed in Chapter 2)
- value (unit cost of the desired outcome)
- optimism bias correction (see section 3.5.4 below).

3.4.2 Benefit calculation[20]

The maximum potential monetary benefit for each outcome is calculated using the following formula (additional technical concepts will be discussed in section 3.5 below):

Figure 8: Benefit calculation[21]



3.4.3 Output

There are a number of ways to present the outcome of the financial measurement and modelling. Table 3 below presents a summary of these.[22]

[19] Models described in proposals and developed during the JDP are necessarily based on hypothesised projections. We envisage that they will be updated with program-generated data over time.

[20] Benefits accruing from social impact investment can often vary over the life of the transaction. It is appropriate for proposals to consider accrual of benefits varying over the life of the transaction. Further refinement of the timing of benefits can be undertaken during JDP.

[21] HM Treasury (2014). *Cost benefit analysis guidance for local partnerships*, section 7.20

[22] We acknowledge that financial modelling is more certain in the immediate timeframe and that benefits might be greater in forward estimate periods.

Table 3: Output of financial modelling

	Definition	Calculation	Application
Net Present Value (NPV)	The difference between the benefits and costs of a program taking into account the differences in the timing of these costs and benefits.	Subtract the discounted costs from the discounted benefits (discount rate is discussed in section 3.5.5).	A positive NPV indicates that the benefits should exceed the costs of a program, which strengthens the proposal.
Benefit Cost Ratio (BCR)	Another way of presenting net present value – this time as a ratio of benefits over costs.	The ratio of discounted benefits over discounted costs.	A BCR >1 indicates that the benefits should exceed the costs of a program, which strengthens the proposal.
Payback period	The timeframe in which the discounted benefit flows from a program begin to exceed costs.	This represents the time-frame in which a program achieves a positive NPV or BCR>1.	The shorter the timeframe, the stronger the case for payment of a dividend on the social impact investment.
Return on investment	This is a restricted version of benefit cost ratio and is measured by assessing the ratio of discounted cost-savings over discounted costs.	Discounted cost savings over discounted costs	The percentage of this ratio above 1 represents the return on investment and is a critical determinant of the dividend payable to investors.

While benefits that accrue to other recipients such as individuals or communities are not included in the financial cost benefit analysis to assess the proposal's viability, they can be discussed qualitatively in the proposal. As noted previously, they could also be quantified as part of a full cost benefit analysis if a proponent has evidence to support these calculations.

3.5 Assumptions, risks and uncertainties

Key issues to consider:

- What assumptions have been made in the calculations?
- How have we accounted for uncertainty?
- How have we accounted for potential variations in the performance of the intervention as it is rolled out across different settings?
- How have we accounted for differences in the timing of costs and benefits?

3.5.1 Monetisation

Monetising benefits can be problematic, particularly when valuing social outcomes from a societal perspective. Such outcomes can be somewhat intangible (e.g. the productivity gains from increased life expectancy or improved educational outcomes) and often come with a wide margin of error. Sensitivity analysis (see below) is one means of addressing this source of uncertainty.

3.5.2 Inflation

Because the appraisal will need to assess costs over a number of years, inflation must be included in the financial model. This involves adjusting for inflation those costs incurred in previous or future

years. In doing so, the analysis can proceed by comparing costs in commensurate real terms. This process is separate, and additional to, discounting (see Section 3.5.5 below).

3.5.3 Optimism bias

Where the providers of an intervention are involved in the analysis, there is the potential to overstate benefits and understate costs. This is known as optimism bias. While it is an issue that can arise during any aspect of evaluation, it is a particular risk in financial analyses due to the assumptions and extrapolations that need to be made (such as those highlighted in this section). Specifying the approach to analysis beforehand and conducting sensitivity analysis (see below) can help mitigate this bias.

Table 4 below provides some guidance to correct optimism bias.

Table 4: Confidence grade for cost data[23]

Confidence grade	Colour coding	Data source	Age of data	Known data error	Optimism bias correction
1		Independently audited cost data	Current (< 1 year old)	+/- 2%	0%
2		Formal service delivery contract costs	1-2 years old	+/- 5%	+ 5%
3		Practitioner monitored costs	2-3 years old	+/- 10%	+10%
4		Costs developed from ready reckoners	3-4 years old	+/- 15%	+15%
5			4-5 years old	+/- 20%	+25%
6		Uncorroborated expert judgement	> 5 years old	+/- 25%	+ 40%

3.5.4 Uncertainties – sensitivity analysis

Sensitivity analysis is defined as testing the sensitivity of results to provide information on the robustness of an intervention to adverse movements in the range of variables determining its viability. The purpose is to indicate the generalisability of the findings to different situations. In social impact investment proposals, at least three performance scenarios need to be considered (i.e. baseline, below baseline, and above baseline scenarios) and should be included in your proposal.

[23] HM Treasury (2014). *Cost benefit analysis guidance for local partnerships*, section 7.20

3.5.5 Discounting

Discounting is an adjustment made to the value of costs and outcomes occurring in the future and is required in financial modelling for proposals. Both costs and outcomes should be discounted, for both the intervention and the counterfactual.

Discounting can be contentious. For example, discounting outcomes can be perceived to downplay the benefits of preventive interventions that occur in the future. While we acknowledge there are arguments for and against discounting, the standard discount rate in NSW is 7% on costs and benefits and is recommended for social impact investment proposals. Discounted values are presented in present values.

3.6 Other financial measurement methods

There are alternative methods for measuring the economic and financial outcomes of social programs. They have been used extensively in a range of sectors, notably health. For the purposes of determining the dividend from a social impact investment, they may be *used only to complement a financial cost benefit analysis*. These are briefly described in Table 5 below.

There are also a number of useful guides for measuring the economic outcomes of an intervention. In particular, proponents are referred to guidance provided by the NSW Government.[24]

[24] NSW Treasury (2017), NSW Government Guide to Cost-Benefit Analysis, TPP 17-03

Table 5: Other measurement methods

Method	Description	Strength	Weakness
Cost-minimisation analysis	When the comparison involves two or more interventions (usually including a status quo option) in which the outcomes are assumed to be or have been demonstrated to be equivalent and thus the comparison is made on the basis solely of cost.	Simple, as the focus is on costs, there is no need to address the uncertainties associated with measurement and valuation of outcomes.	A narrow form of assessment; the assumption of equivalent outcomes is often difficult to justify.
Cost-effectiveness analysis	When interventions being compared are similar to the extent that their outcomes can be valued in the same units. Produces an incremental cost-effectiveness ratio presented in terms of a cost per unit of outcome gained relative to the comparison (e.g. incremental cost per case of reoffending prevented, incremental cost per case of disease prevented, etc.).	Provides a transparent means of comparing the costs and outcomes of interventions.	Potential weakness is the comparability of the relative value of an outcome across different contexts, making it difficult for a decision maker to benchmark in deciding what constitutes value for money.
Cost-efficiency analysis	Compares options in terms of cost relative to a common measure of output (e.g. incremental cost per case treated, client visited, service delivered, procedure performed etc.).	Enables individual organisational units (such as hospitals and schools) to be assessed in terms of organisational efficiency.	Focus on service outputs rather than outcomes/impact. These methods are generally used when a decision has been made to implement an intervention to achieve particular outcomes.
Cost-utility analysis	Cost-utility analysis is a tool developed by economists for the purposes of evaluating health sector programs. Use either Quality Adjusted Life Years (QALYs) or Disability Adjusted Life Years (DALYs) as outcome measures. They are recommended for use in guidelines for health regulatory assessments, such as those produced by the Pharmaceutical Benefits Advisory Committee in Australia and the National Institute for Health and Clinical Excellence in the UK.	QALYs or DALYs as outcomes can be employed as a means of comparing across diverse sets of programs.	
Cost-benefit analysis based on social return on investment (SROI)	SROI is an approach to assign a monetary value to the social, economic and environmental outcomes created by an activity or an organisation. It is based on a set of principles that are applied within a framework.	Provides a societal perspective and helps distinguish those programs that are genuinely cost-saving from those that merely shift costs from government to other sections of the community	In principle includes intangible outcomes that can be difficult to quantify.

4. References

Austin P (2011). *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies*. Multivariate Behav Res. 2011 May; 46(3): 399–424

Cave S, Williams T, Jolliffe D, Hedderman C (2012). *Peterborough Social Impact Bond: an independent assessment Development of the PSM methodology*. Ministry of Justice Research Series 8/12 May 2012 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217392/peterborough-social-impact-bond-assessment.pdf

Cohen, J (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. NY:Lawrence Erlbaum

Flatau P, Zaretzky K, Adams S, Horton A, Smith J (2015). *Measuring Outcomes for Impact in the Community Sector in Western Australia*, Bankwest Foundation Social Impact Series, Issue 1, March 2015

Haynes, L., Service, O., Goldacre, B. & Togerson, D. (2012). *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London: Cabinet Office

Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ (2015). *The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting*. BMJ. 2015 Feb 6;350:h391. doi: 10.1136/bmj.h391

HM Treasury (2014). *Cost benefit analysis guidance for local partnerships*

HM Treasury (2013). *Green Book Supplementary Guidance on Public Sector Business Cases Using the Five Case Model*, Lowe, HM Treasury; NSW Treasury guidance

HM Treasury (2011). *The Magenta Book. Guidance for evaluation*. April 2011. Accessed at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

Ivy So and Adam Jagelewski (2013). *Social Impact Bond Technical Guide for Service Providers*. MaRS Centre for Impact investing. November 2013

Jolliffe D, Hedderman C (2014). *Peterborough Social Impact Bond: Final Report on Cohort 1 Analysis*. Prepared for Ministry of Justice 7th August 2014 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/341684/peterborough-social-impact-bond-report.pdf

Kellogg, W. (2004). *Logic model development guide*. Michigan: WK Kellogg Foundation

Morris, S. et al. (2014). *Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis*. BMJ, 349, g4757

Muir, K and Bennet, S (2014). *The Compass: Your guide to social impact investment*. Centre for Social Impact Assessment: Sydney

NSW Government (2015). *Principles for social impact investment proposals to the NSW Government*

http://www.dpc.nsw.gov.au/__data/assets/pdf_file/0006/171897/Social_Impact_Investment_Proposal_Principles.pdf

NSW Government (2015). *Social Impact Investment Policy* http://www.dpc.nsw.gov.au/__data/assets/pdf_file/0011/168338/Social_Impact_Investment_Policy_WEB.pdf

NSW Treasury (2017). *NSW Government Guide to Cost-Benefit Analysis*, TPP 17-03

https://arp.nsw.gov.au/sites/default/files/TPP17-03_NSW_Government_Guide_to_Cost-Benefit_Analysis_0.pdf

Shahidur R. Khandker Gayatri B. Koolwal Hussain A. Samad (2010). *Handbook on Impact Evaluation*. The World Bank: Washington D.C

Social Finance (2013). *A Guide to Social Impact Bond Develo**o**pment*. Social Finance: London (www.socialfinance.org.uk)

Social Finance (2015). *Technical Guide: Designing outcome metrics*. Social Finance: London

Weatherburn, D (2009). *Policy and program evaluation: Recommendations for criminal justice policy analysts and advisors*. Sydney, N.S.W.: Bureau of Crime Statistics and Research New South Wales, 2009. Crime and justice bulletin: contemporary issues in crime and justice; ISSN:1030-1046; no. 133

5. Tables and Figures

Tables

Table 1: Summary of the different categories of costs 33

Table 2: Different benefits and how they contribute to the complexity of social impact investment transaction 35

Table 3: Output of financial modelling 37

Table 4: Confidence grade for cost data 38

Table 5: Other measurement methods 40

Figures

Figure 1: When to use this guide? 9

Figure 2: Hierarchy of scientific evidence 15

Figure 3: Power and sample size 16

Figure 4: Program logic examples 17

Figure 5 Hierarchy of Evidence Quality for Counterfactual Design Options 24

Figure 6: Different benefits and how they contribute to the complexity of social impact investment transactions 30

Figure 7: Visualisation of social impact bond financial model 31

Figure 8: Benefit calculation 36