



**GOVERNMENT  
OUTCOMES  
LAB**

**HOW TO  
GUIDE**

---

**SETTING AND  
MEASURING  
OUTCOMES**

**A guide to  
identifying, defining  
and measuring  
outcomes for the  
purposes of outcome  
based commissioning**

# Contents

<b>About the Government Outcomes Lab .....</b>	<b>3</b>
<b>About the guides .....</b>	<b>4</b>
<b>General guidance on outcomes and outcome measures .....</b>	<b>5</b>
Introduction and overview .....	6
Understanding outcomes, measures and metrics .....	7
Different types of outcome .....	8
Individual versus cohort measurement .....	10
Developing an outcomes framework .....	12
Setting outcome metrics and triggers .....	15
Mitigating the impact of uncontrolled external factors .....	16
Avoiding or mitigating perverse incentives .....	17
Involving providers and investors .....	18
Evaluation .....	19
<b>Toolkit: Setting and measuring outcomes .....</b>	<b>20</b>
Feasibility .....	21
Setting and measuring outcomes: summary of key actions .....	21
1. Confirm the availability and measurability of outcomes .....	22
2. Estimate the costs incurred by failure to achieve the outcomes .....	22
<b>Development .....</b>	<b>23</b>
1. Define the outcomes and how best they can be measured .....	23
2. Specify metrics and triggers that will drive payment under the contract .....	24
3. Develop a payment structure that sets out the payments to be made .....	25
4. Engaging and consulting providers and investors .....	26
<b>Implementation .....</b>	<b>27</b>
1. Selecting the appropriate procurement procedure .....	27
2. Defining outcomes in the specification .....	28
<b>Operation .....</b>	<b>29</b>
1. Reviewing the outcomes framework and the measures and metrics in use .....	29
<b>Evaluation .....</b>	<b>30</b>
1. Evaluation of the outcomes framework and payment structure/tariff .....	30



## About the Government Outcomes Lab

The Government Outcomes (GO) Lab at the Blavatnik School of Government was founded in 2016 as a new centre of academic excellence for innovative government commissioning. Through rigorous academic research, the GO Lab aims to deepen the understanding of outcome based commissioning and to provide independent support, data and evidence on the effectiveness of the outcome based commissioning model. To achieve this, the GO Lab works with a wide range of stakeholders and partners from across the public, voluntary and private sectors.

### Our mission

Outcome based commissioning has emerged as an innovative way for governments around the world to achieve better social outcomes for the most vulnerable people in society. While numbers of, and funding for, outcome based approaches have increased over recent years, research has not kept pace with this speed of

growth. Much is still unknown about whether outcome based commissioning will meet its promise. The GO Lab harnesses expertise from across Oxford University and other partners to enhance the understanding and existing research on outcome based commissioning. It also builds on the evidence base to evaluate the effectiveness of this model versus alternatives and to support local authorities that are developing and implementing an outcome based approach.

### Thanks and acknowledgements

This guide has been written by Neil Stanworth from [ATQ Consultants](#). We are indebted to him for sharing his very considerable expertise and experience for the wider benefit of the commissioners and providers of outcome based contracts in this guide.

We would also like to thank Roger Bullen from [Triodos Bank](#) and Jen Byrne from [Future Public](#) for reviewing and commenting on the drafts.



## About the guides

### Beta status

This guide is one of a series we are publishing during 2017. The purpose is to provide guidance to commissioners and is based upon the best advice of practitioners and experts with experience in outcome based commissioning at the point of writing. It is an emergent area and we know practice will evolve rapidly. We have designated this and other guides as ‘beta’ documents and we warmly welcome any comments or corrections. We will regularly review and update our material in response. Please email your feedback to [golab@bsg.ox.ac.uk](mailto:golab@bsg.ox.ac.uk)

### Scope of the guidance

This guide aims to provide advice on how to go about identifying, defining and measuring

outcomes when putting in place an outcomes-based contract or social impact bond. It is aimed mainly at local commissioners, but should also be useful to providers of services and interventions who are involved in the development of an outcomes-based contract – especially when the contract is ‘provider-led’.

### Structure

The guidance is in two parts. The first is general guidance on the different types of outcomes and measures available, and issues that need to be thought about in setting appropriate outcomes and measures. The second provides a checklist of key actions and issues to consider at each stage of the process of developing and implementing an outcomes-based contract. The parts can be read separately but are best read together.

# GENERAL GUIDANCE ON OUTCOMES AND OUTCOME MEASURES

## **|** Included in this section

Introduction and overview

Understanding outcomes, measures and metrics

Different types of outcome

Individual versus cohort measurement

Developing an outcomes framework

Setting outcome metrics and triggers

Mitigating the impact of uncontrolled external factors

Avoiding or mitigating perverse incentives

Involving providers and investors

# Introduction and overview

By definition, setting the right outcomes, and measuring them effectively, is one of the most important aspects of an outcomes-based contract.

Getting this right is even more important when it is a payment by results (PbR) contract, or a social impact bond (SIB), because the achievement of at least some of the outcomes will determine how much is paid by commissioners to providers and/or investors, and the definition of outcomes directly affects whether commissioners achieve value for money from the contract.

In an outcomes contract, providers – along with investors – will typically be given more freedom to operate than in forms of contract that pay for defined outputs and prescribe an expected method of delivery. In payment by results and SIB contracts, the level of prescription is less and commissioners need to use the measurement framework to ensure that providers deliver a service acceptable to all the constituencies and employ acceptable means of achieving outcomes, as well as achieving the outcomes themselves.

This guide aims to help commissioners make good decisions about the choice of outcomes and associated measures to apply in a contract, and set sensible metrics or defined levels against which performance can be measured and payments made.

It is important to note that in considering and setting outcomes there are no hard and fast rules about what constitutes the ‘right’ set of outcomes and measures (often termed an outcomes framework) for a particular contract. In most cases, the task of setting outcomes is likely to proceed incrementally across the project development phase and will be informed and influenced by other decisions, in particular the design of the payment mechanism.

Different stakeholders may have different perspectives on what constitutes a robust and viable outcome measure. Project teams should expect and accept that the process of developing outcomes will be both:

- **iterative** – because outcomes and measures will be identified but may then be discarded (for example because of the challenges of data collection). There may therefore be an element of trial and error involved; and
- **progressive** – because in the early stages of development (and especially when assessing the feasibility of the contract) outcomes and measures need not be specified and defined in great detail, but as work progresses into development and implementation, outcomes measures and metrics will need to be more clearly and closely defined, and ultimately specified tightly and unambiguously in contractual terms.

This section is about creating a framework for real-time measurement and evaluation with the primary (although not exclusive) intent of enabling payment by results. Many projects, particularly those in receipt of co-investment from government or the Big Lottery Fund, will also invest in independent evaluation. In some cases the results of this evaluation is linked to the system of reward, but most often not in the UK. In general terms, there is a linear relationship between the confidence that commissioners have in the impact a service has had and the value it has created and the stage at which the measure is applied. In most payment by results contracts there will be a compromise on this quality of confidence at the point of payment in order to make the contract affordable and fundable.

This first chapter provides general guidance on outcomes setting and the terminology that

surrounds them, irrespective of where the commissioner is in the development cycle. Chapter 2 then provides further and more

detailed guidance on key steps to developing and managing an outcomes framework at each stage of the development process.

## Understanding outcomes, measures and metrics

Many of those involved in setting outcomes, and much of the guidance around them, tends to use the terms *outcome*, *measure* and *metric* loosely and interchangeably. They do however mean slightly different things, and should really be seen as separate parts of the process of developing an outcomes framework and associated contract:

- an **outcome** is what changes for an individual (or other defined unit within, such as a family) as the result of a service or intervention (e.g. improved learning outcomes);
- an **outcome measure** (also termed an indicator) is the specific way the commissioner chooses to determine whether that outcome has been achieved (e.g. a test score). Often this encompasses a single dimension of an outcome; and
- an **outcome metric** (also termed trigger or target) is the specific value attached to the measure for the purposes of determining whether satisfactory performance has been achieved (e.g. a test score of 95 out of 100 or improvement of 30 points in a test score over a 5 month period). Under a PbR contract or SIB, these metrics will usually determine whether a payment is made to the provider or investor.

Both measures and metrics may relate to what is achieved by and for an individual within the cohort, or to performance across the cohort as a whole. See ‘individual versus cohort measurement’ below for further guidance on this.

It is helpful to think about outcomes and measures separately and sequentially in thinking through the framework as a whole. It may be possible to define outcomes that relate fully and holistically to the policy objective, but measures may not be comprehensive in capturing all facets of an outcome. The fact that an outcome is difficult to measure does not make the outcome invalid, but it may mean a PBR approach is not appropriate.

It is also advisable to have a fairly settled view of the outcomes and measures to be applied to a contract before starting to assign specific values to them as metrics or triggers.

To take some simplified examples:

For a contract that aims to reduce the number of young people who are Not in Employment, Education or Training (NEET):

- a key outcome might be that the young person is in employment;
- the outcome measure could be confirmation from the employer that the person is employed by them; and
- the outcome metric (and payment trigger) could be that the young person is in continuous employment of a minimum of 16 hours per week for a defined period or that 20% of the total cohort are in continuous employment for a defined period on average.

For a contract that aims to improve the emotional wellbeing of young people:

- the outcome is improved emotional wellbeing of young people
- the outcome measure could be that there is an identifiable improvement in young

- people's resilience and ability to deal with challenges using the Strengths and Difficulties Questionnaire (SDQ); and
- the outcome metric (and payment trigger) could be that the young person reduces their

total SDQ score by a defined number over a specified period or that there is a mean reduction in the average score across the cohort as a whole

## Different types of outcome

In the terminology that is often used to describe outcomes, they can be divided into:

- hard outcomes**, which can be objectively and independently measured
- binary outcomes** – a type of hard outcome that has only two states – either it is achieved or it is not

- soft outcomes** – which depend on subjective assessment.

These types of outcome and their advantages and disadvantages are summarised in Table 1.

Type of outcome	Definition	Advantages	Disadvantages	Simple example
<b>Hard outcome</b>	Any outcome that can be measured objectively	Usually simpler to measure. No risk of disagreement about outcome achievement	Not always available. May not capture sustained impact if used in isolation. May not reflect those matters most important to service users or practitioners	Achievement of specific education or skills qualification
<b>Binary outcome</b>	Hard outcome which has only two states- achieved or not achieved	As above with even less likelihood of disagreement	As above. Risk of 'parking' if the outcome can no longer be achieved	In employment
<b>Soft outcome</b>	Outcome that requires subjective assessment of its achievement – usually applied to an individual	Useful when a hard outcome is not available or suitable. Also used to test progress towards a hard outcome. Measures whether the service meets the expectations of service users	Consistency of measurement can be difficult. Potential for disagreement about achievement. Potential for gaming	Improvement in score using recognised scale and assessment tool – e.g. Strengths and Difficulties Questionnaire

Table 1: Outcome measures – advantages and disadvantages

**Hard outcome measures** are useful (and often preferred in outcomes-based contracts) because they leave little room for doubt or disagreement as to their achievement, and are

usually easy to measure. There are numerous examples of hard outcome measures, including a child being in or out of care, an adult being in employment, a young person being in

education etc. Hard outcome measures can often be underpinned or validated by a statutory process – for example there can be no doubt about a child being on a Child Protection Plan, because being on such a Plan requires a formal decision of the local authority. They can also relate to easily measurable objective milestones such as achievement of a defined education or skills qualification.

The Ways to Wellness SIB, commissioned by Newcastle and Gateshead CCG, measures improvements in the management of long term conditions achieved through social prescription. It uses a combination of:

- A hard outcome measure, relating to reduced hospital admissions and use of outpatient & A&E services; and
- A soft measure of improved well-being, measured through Triangle Consulting's Wellbeing Star.

The main disadvantage of hard outcome measures is that, used in isolation, they can give an incomplete picture of the value of the outcome to the individual, or bias the measurement towards outcomes that are not significant to the experience of the individual and especially may not capture the sustained impact of the service. For example, a young person may re-engage in education, but without resolution of underlying issues may disengage after a short period.

**Binary outcome measures** are sometimes confused with hard outcome measures but are not quite the same thing. For example, reduced admission to hospital is a hard outcome, but only a complete absence of admission (or not) is a binary outcome. Binary outcome measures have the same advantages as all hard outcomes in being unarguable and easy to measure, but similar disadvantages, with increased risk that they can lead to perverse incentives.

In particular, they can increase the risk that service recipients can cease to receive help – sometimes known as ‘parking’ – because the binary outcome can no longer be achieved. For example, a binary outcome that the individual achieves complete desistance from substance misuse can no longer be achieved if there is a single instance of misuse during the measurement period, potentially discouraging further intervention with that person. Binary outcomes have been used where it has been deemed unacceptable for the public sector to pay for outcomes that include negative events, for example, payment for a reduction in re-offending may still be made when a level of offending that some might consider unacceptable has taken place. Commissioners should consider the political impact of the measurement framework in the process of designing the measures and metrics.

**Soft outcome measures** are nearly always measured at the level of an individual or family unit, rather than across a cohort as a whole. Examples include the quality of wellbeing experienced by an individual, the capacity of a family to manage a challenging child, or the ability of a person to look for work. Soft outcomes are usually set and measured by reference to proprietary measurement methods, such as SDQ or the Outcomes Star family of tools.

Soft outcome measures are useful in a number of contexts, including i) where a hard outcome measure is not available or is impractical to use, ii) to indicate progress towards a hard outcome that may take a long time to achieve, or iii) to reinforce hard outcome measures that in isolation, may give a misleading picture of the sustainability of an outcome. For example:

- many health outcomes (such as improved emotional health and wellbeing) are very difficult to measure objectively, and a soft outcome is nearly always used;

- for a disengaged young person who is homeless, moving into settled accommodation is a hard outcome measure, but may take a long time to achieve. A soft outcome related to the improved resilience and capacity of that person may be a useful shorter-term indicator of the likelihood of them achieving that outcome in due course, as well as capturing whether the provider is offering suitable accommodation options in the view of the service user; and
- the reunification of a child who has been in care with its family is a hard outcome

measure that can be reinforced by assessing the capacity and resilience of the family, the child, or both.

The major disadvantage of soft outcome measures is that, because they are measured subjectively, there is scope of disagreement about the level of achievement. There is also increased risk of results being skewed (or even deliberately distorted) by the way the assessment is carried out. This needs to be taken into account in the way the outcome and its measurement is defined and the process of measurement is specified.

## Individual versus cohort measurement

The basic principle behind outcomes based commissioning and contracting is that a group has been identified which is experiencing adverse outcomes, and those outcomes can be changed (or eliminated) by a service or intervention. The success of the contract in achieving such change (and triggering payment where appropriate) can be measured in two different ways. These are by reference to:

- the outcomes achieved by an individual (or sometimes by a unit such as a family); or
- the total outcomes achieved across the target cohort, expressed either as a total number of changed outcomes or, more usually, a percentage change in total or average outcome across the cohort.

Both individual and cross-cohort outcomes can be expressed as either a positive (the outcome is achieved, or there is percentage improvement in the outcome) or a negative (the adverse outcome is eliminated, or there is a percentage reduction in the outcome).

One reason why it is advisable to separate identifying the outcome itself from its measurement, is that individual and collective

measurement have different strengths and weaknesses, and lead to a different payment structure if the contract involves PbR or is a SIB.

### Individual outcome measurement

- works best when the cohort is relatively homogeneous, and all those within the cohort are experiencing or causing the same or similar adverse outcomes (or are predicted to do so). This means that outcomes (and how they change) can safely be set at an individual level. An example would be avoiding entry to care for a cohort all at imminent risk of entering care;
- does not involve a comparison group or other baseline against which success can be measured for the purposes of payment. Instead, payments relate to the changed/improved outcomes per individual in the form of a tariff or rate card which specifies what will be paid per individual per outcome;
- allows for 'deadweight' within the payment level set. Since there is no comparison group the outcomes achieved are likely to include some that would have happened without the intervention. This means that the payment

has to be set at a level which will ensure the provider (and investor) is incentivised to achieve additional outcomes. This can only be avoided if the commissioner is certain that the intervention is being applied only to people who would not otherwise achieve an improved outcome; and

- requires good evidence of the likely level of deadweight. It follows that the commissioner needs to know how many outcomes would have been achieved without the intervention. If not, there is significant risk that the outcome payments will not be at the right level and that the provider/investor will be rewarded for achieving too little impact. There is also a risk that there will be much less additionality in the contract than the commissioner need or expects.

## Cohort outcome measurement

- works best if the current adverse outcomes vary considerably across the cohort, making it more difficult to set standard measures of success at an individual level. An example is offending, where both the frequency and severity of offences committed (as measured by convictions) can vary widely;
- usually requires the level of outcomes achieved through the contracted intervention to be compared against a comparator group who did not receive the service, i.e. what would have happened without the intervention, for the purposes of assessing whether a payment should be made. The comparator can take a number of forms, including a historic baseline, a modelled prediction of the future baseline, a statistical comparison group or a control group. Each of these approaches has benefits and disadvantages which are explored in the GO Lab's *An Introduction to Evaluation* guide; and
- does not normally require a separate calculation of deadweight – because

any deadweight will be reflected in the measurement against the comparator group, and only performance over and above the comparator group will be rewarded.

Experience to date shows that the majority of outcome-based contracts in the UK have been let using a tariff or rate card linked to individual outcomes (including nearly all such contracts put in place by central government), and a growing proportion of contracts let by local commissioners. This is largely because the tariff-based approach is cheaper to manage and usually allows payments to be made more quickly and regularly than a cohort-level outcome measurement approach that requires data on the whole cohort to be collected before payment can be made.

Cohort-level measurement does however have advantages as described above. It was the basis of measurement in some early SIBs in the UK such as the Peterborough One SIB, and is used extensively in Pay for Success Bonds in the USA.

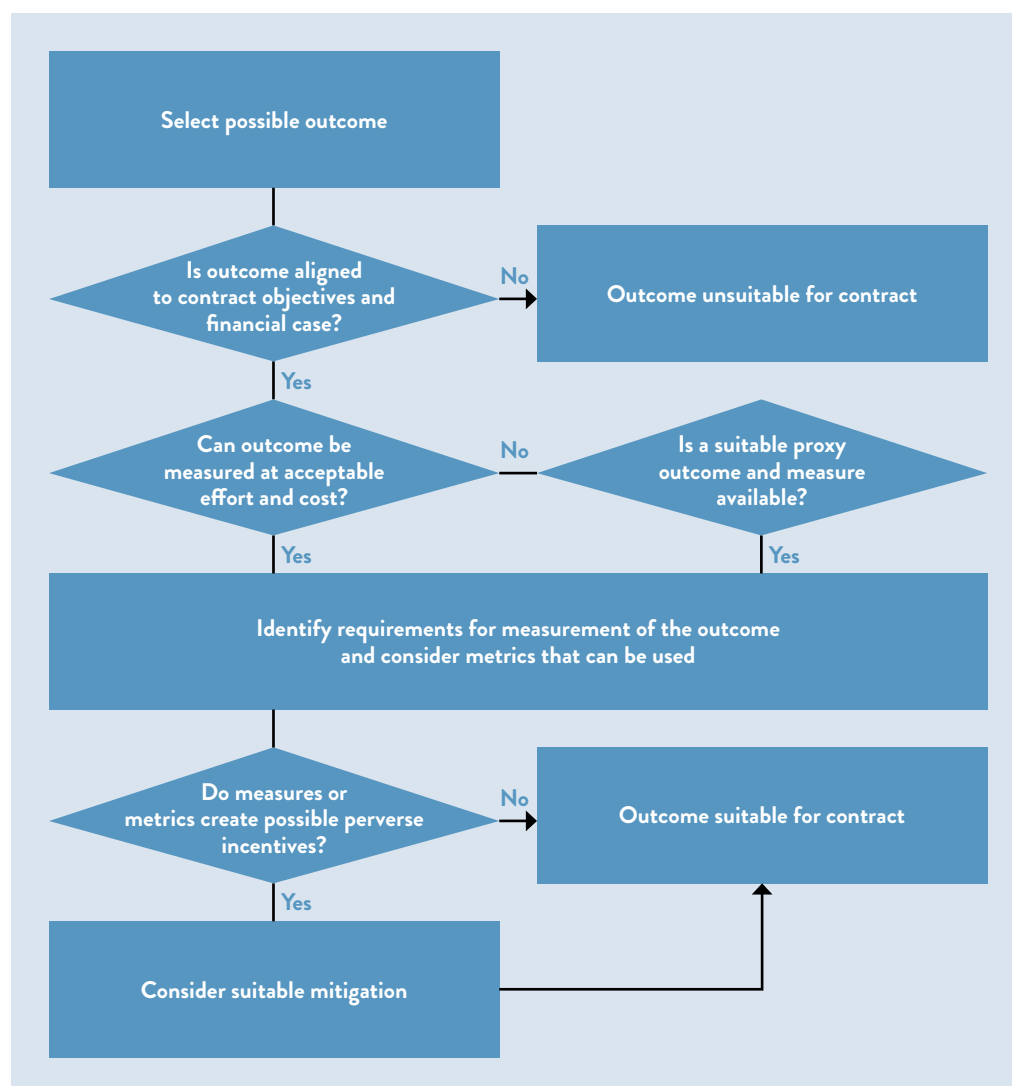
In developing an outcomes framework, therefore, the pros and cons of individual versus cohort measurement need to be carefully considered. In particular, commissioners should consider using cohort-based measurement if:

- the prevalence of current adverse outcomes or needs varies considerably across the cohort, making it more difficult to set a flat tariff or rate card for all participants; and/or
- the commissioner cannot predict the level of deadweight with reasonable accuracy, and therefore cannot build an estimate of deadweight into the rate card.

# Developing an outcomes framework

In testing the feasibility of a SIB, and then developing and implementing a SIB that proves to be feasible, commissioners should develop an outcomes framework which defines:

- the outcomes to be used;
- the measures to be applied to each outcome;
- the specific metrics to be applied to each measure, that determine outcome payments to be made; and
- when measurement takes place



Identifying outcomes – simplified decision tree

## Identifying the right outcomes

Commissioners should aim to identify those outcomes that:

- **align to the objectives and critical success factors of the SIB** including:
  - the social problems that the contract aims to address; and/or
  - the financial benefits that it will bring, such as savings to the commissioners or reductions in current service demand;
- **can be measured at reasonable cost and effort** – see *Determining how outcomes will be measured*;
- **are likely to be acceptable to all the key parties to the SIB**. This should be tested

during detailed development – see *Involving providers and investors*;

- **can be mainly attributed to the intervention** and are not unduly susceptible to uncontrolled external factors (sometimes termed exogenous factors) – see *Mitigating the impact of uncontrolled external factors*;
- **reflect the priorities of service users (are co-produced) and indicate quality of delivery**; and
- **avoid or minimise perverse incentives** – see *Avoiding or mitigating perverse incentives*.

Relatively few outcome measures will perfectly capture the overarching outcome intent of a programme. Most measures will have drawbacks in one or more of these areas, and commissioners should consider carefully the benefits of an outcome measure against its disadvantages. In general, if the outcome cannot be measured it should not be used unless a suitable proxy outcome and measure is available – see *Use of proxy outcomes and measures*. However, if an outcome measure aligns well to the objectives of the contract, commissioners should consider whether:

- any investment required to collect outcome data is justified by its other benefits;
- the impact of external factors can be accounted for in the measurement process or the metrics set; and/or
- the risk of perverse incentives can be mitigated.

## Primary, secondary and other outcomes

In developing an outcomes framework it is common to identify both primary and secondary outcomes, where:

- the **primary outcome** measure is the single most important outcome from the contract, and the one that the commissioner most wants to see positively impacted; and
- **secondary (or additional) outcome measures** are the other important outcomes that

the commissioner wishes to see improved.

They may capture a different dimension of the programme intent or may serve to counterbalance perverse incentives, or ensure that service users are satisfied with services.

The hierarchy of primary and secondary outcomes will usually be reflected in the payment structure (whether set as a tariff or based on cohort measurement) with the majority of payment being attached to achievement of the primary outcome, and lesser payments attaching to secondary outcomes. This ensures that the incentives to provider(s) and investor(s) are aligned to the priorities of the commissioner(s).

It is advisable to limit and classify the outcomes in this way because fewer outcomes are simpler to manage, and because multiple outcomes make it harder to predict when outcomes will occur – and thus forecast commissioner expenditure, and income and cashflow for the provider (and investor if appropriate). It is a requirement of both the Life Chances Fund (LCF) and its predecessor the Commissioning Better Outcomes (CBO) Fund, that applicants for funding identify a single primary outcome and a limited number of secondary outcomes. The LCF allows for up to five secondary outcomes, but we recommend keeping the number of outcomes linked to payment fewer than this (usually no more than 2–3) to avoid overcomplicating contract management.

Some SIBs may have only one outcome measure, to which all the payment is linked, and there is no requirement for multiple outcomes to be set if they are not needed.

Typically, the primary outcome measure will reflect either or both:

- the most important outcome for the commissioner in policy terms; and/or
- the outcome that delivers the greatest financial benefit to the commissioner.

The choice of the right secondary outcomes may be less straightforward but will typically involve one or more outcomes that:

- are important to one or more commissioners in policy terms; and/or
- reinforce and sustain the primary outcome
- capture stakeholder interests not reflected in the primary outcomes.
- offer further significant financial benefit to the main or other commissioners;
- counteract tendencies to ‘game’ the primary outcome (by reducing incentives for creaming and parking or delivery of services that are otherwise detrimental/partial in their support).

In some contracts, the commissioner may wish to measure a larger number of outcomes than the norm. In these cases, we recommend defining and specifying outcomes that will be measured but will not be linked to payment. Commissioners should however take account of the cost and effort required to collect and monitor data on any further outcomes – both to themselves and to providers and investors.

## Determining how outcomes will be measured

### General considerations

Commissioners should consider carefully how an outcome will be measured and how the data needed to support each measure will be collected, reviewed and reported. Among the factors to be considered are:

- whether the outcome is best measured using an individual or collective (cross-cohort) approach – see *Individual versus cohort-level measurement*;
- whether the data needed to measure the outcome are already collected for another purpose, for example government statistical returns or internal performance management;
- if not, whether the data collection requires

significant investment in new collection processes and systems;

- who will be responsible for collecting the data and whether they have the capacity to do so; and
- whether the data need to be independently checked and validated.

Commissioners should be careful not to make assumptions about the availability of data from other parties or the ability of those parties to collect data on their behalf. If such data are needed to measure an outcome, its availability and the legality and practicality of its collection should always be confirmed with the relevant collection agency. An early audit of data during the feasibility stage, will help inform the design of the outcomes framework.

### Use of proxy outcomes and measures

If the primary or other important outcome of a contract cannot be easily measured, it may be possible to use a proxy measure that reliably indicates whether the desired outcome has been achieved. These are usually known as proxy (or surrogate) measures. The Centre for Government (US) defines a proxy as an “indirect measure of the desired outcome which is itself strongly correlated to that outcome used when direct measures of the outcome are unobservable and/or unavailable”.

A good example of a proxy outcome measure is proven offending (as measured by prosecution leading to a conviction) which is widely used as a proxy for all offending.

The reason conviction is a proxy measure is because convictions only account for offenses which make it into the justice system. The fact is, many offences go undetected and thus total offending cannot reliably be measured. Nevertheless, the two are closely linked. Many outcomes which indicate a reduction in demand (such as reduced hospital admissions) are proxy measures of the true outcome – in this case improved health of the individual.

Many soft outcomes and measures are used as a form of proxy, and on the basis that they will indicate achievement of another outcome that is difficult to measure or that will give an unreliable indication of the outcome – for example, a soft measure of improved resilience can be used as a proxy for care leavers to secure stable independent living.

Importantly, the risk in using proxy outcomes is that strong correlation does not mean causation: you may end up paying for outcomes which have little to no effect on the actual policy intent of the social impact bond.

### **Lead or progression outcome measures**

In some cases it will be desirable to define measures that indicate progression towards the achievement of the primary outcomes because the period of measurement of the primary outcomes is lengthy and the commissioner or other parties will want reassurance that the primary outcomes are likely to be achieved. It can also be the case that there is a very significant risk associated with achieving the primary outcomes for the whole cohort

because there is a variance in presenting need and therefore a different starting point in achieving the primary or secondary outcomes, or an acceptance that only some of the cohort will achieve the outcomes. To avoid the risk of “parking”, the commissioner may wish to incentivise the provider to continue to invest in the whole cohort by paying a differential rate for different parts of the cohort and/or paying for measures (often termed lead or progression measures or milestones) that suggest valuable progression. For example, for children with highly complex needs in foster care, a lead measure might be better attendance at school, where the evidence indicates this increases the likelihood of the child maintaining a stable foster care outcome and there is value and impact for the individual and the service from that outcome. The main disadvantage of such measures is that providers may not invest in moving the cohort towards the primary outcomes if the payment terms provide significant reward at this earlier stage. The level of payment attached to such outcomes therefore needs to be set with care.

## **Setting outcome metrics and triggers**

In addition to identifying and defining outcome measures, commissioners will need to assign specific levels of performance (sometimes known as metrics or triggers) against which success will be judged and, in a PbR or a SIB contract, outcome payments will be made.

Provided the right measures of outcome have been put in place, the setting of appropriate metrics should be relatively straightforward, and should, like measures themselves, align mainly to the social objectives of the contract and to the anticipated financial benefits. An important additional consideration, however, is that metrics should take account of the existing evidence for the effectiveness of the

intervention that is being commissioned and the effect of “deadweight” e.g. the number of job seekers who would return to work anyway without the intervention. Commissioners may wish to set stretch targets for providers under a contract, but it is inadvisable – and unlikely to lead to a successful contract – if the performance required significantly exceeds the available evidence of what can reasonably be achieved. Where a commissioner is uncertain whether the performance level they are requiring is reasonable and attainable, they should test their thinking with providers and investors during the development and implementation stages of the contract.

In some cases, the outcome metric will be the simple achievement of a hard outcome (especially if it is a binary outcome measure – see *Different types of outcome*). In most cases, however, commissioners should consider whether to set a number of additional outcome metrics that reward:

- progression towards the main outcome (sometimes called Lead, Progression or Intermediate outcomes – see above); and/or
- the sustainment of the outcome and/or further improvement by the service recipient.

For example, in a contract that incentivises skills attainment by young people, there may be a metric (and payment) relating to:

- the young person engaging in skills training (a progression or intermediate metric);
- the achievement of a first Level 3 qualification (the main outcome metric); and
- the achievement of a second, level 4 qualification (a further improvement metric).

In a contract that rewards the achievement of employment there may be similar metrics relating to the person:

HCT Group have developed and implemented a SIB that funds independent travel training for young people with special Educational Needs or Disabilities. Contracts with Lambeth and Norfolk Councils are based on the single easy-to-measure outcome that the young person is no longer using specialist transport provided by the LA and is using public transport to get to school or college. This outcome has three payment triggers that reward the achievement of the outcome and its sustainment for two school terms.

- engaging with the programme (a lead metric)
- being in full-time employment for 13 weeks (the main outcome metric); and
- sustaining full-time employment for 26 weeks (sustainment metric).

In setting time-related sustainment metrics, commissioners should be careful to avoid creating perverse incentives – see *Avoiding or mitigating perverse incentives*.

## Mitigating the impact of uncontrolled external factors

In setting outcome measures it is advisable for commissioners to avoid measures that can be unduly influenced by factors outside their or other parties' control. While it is arguably impossible to identify social outcomes that are entirely impervious to wider changes in social policy, or to factors such as demographic or labour market change, it is wise to avoid an outcome that can easily be under- or over-achieved due to external factors such as the decisions of people who are not parties to the contract. Examples include measures that are highly susceptible to:

- possible changes in central government policy – for example an outcome relating to educational attainment might be affected by changes in the national curriculum and examination system;
- planned service changes – for example reorganisation of a Children's Services Department which might affect outcomes from an intervention focused on children in need; and
- changes in local practice – for example the validity of a comparison group (or deadweight assumption) might be undermined by plans

by another body commissioning a similar intervention aimed at all or part of the same cohort or in the same area.

Accepting that with complex social problems it will rarely be the case that a new service is impervious to the influence of a wider system of service, it is important that these influencing factors are made explicit in the process of designing the service and the mechanisms for determining its relative impact. Those influences may be factored in a number of ways, for example by:

- having agreements with other practitioners to operate in a way that enables the outcomes to be achieved; or

- taking account of the likely influence of these factors (positive or negative) in the metrics.

In some circumstances, commissioners may accept obligations under the contract to manage the impact of these factors, but this is often not the case. It may be preferable to include the monitoring and reporting of these impacts as part of the contract management process, with the expectation that the commissioner will take steps to mitigate or avoid where possible.

## Avoiding or mitigating perverse incentives

Perverse incentives are incentives that encourage contract stakeholders to behave in a way that is detrimental to contractual goals even if some outcome metrics improve. The result of perverse incentives can be negative consequences for service users. Examples include measures and metrics that link success to:

- the absence of/reductions in referrals to a statutory agency – which can lead to an incentive not to make a referral when it would be in the best interests of the individual;
- a simplistic binary outcome measure, which can easily be ‘failed’ by a high proportion of the cohort – potentially leading to those that can no longer achieve the outcome being ‘parked’ by the provider. For example, an outcome for a homeless person that can be achieved only if they remain continuously in accommodation for a defined period, will be failed if the person leaves that accommodation for a short period;

- setting a metric that is achieved with varying degrees of effort for different members of the cohort. For example, an employment support programme where some members of the cohort are much further from the labour market than others, may encourage ‘creaming’ (focus of attention on easier to help) and ‘parking’ (neglecting those who are less likely to achieve a positive outcome).

A structuring of the outcome metric that establishes the start point of the individual in terms of need and that rewards the provider according to the distance travelled could serve to reverse the perverse incentive to work with those less vulnerable people where outcomes are more easily achievable;

- a single time-related milestone at which the person receiving the intervention is deemed to have achieved a positive outcome – sometimes termed a ‘cliff edge’. For example, if a successful outcome is defined as a young person remaining out of care six months after intervention, there are potential

perverse incentives for both the provider and the commissioner. The provider might avoid escalation to care until that milestone is reached and payment triggered – even when it is not in the best interests of the young person. Conversely the commissioner might insist that they return to care before the milestone is reached and thereby avoid having to make a payment.

As indicated in the last example, perverse incentives may be created for any or all parties to a contract and not just for providers and investors.

If there is a significant risk of a perverse incentive arising for any party, commissioners should consider how it can best be mitigated:

- any perverse incentives that relate to referral to a statutory body or escalation to a defined status on the decision of one party to the contract can be fairly easily mitigated by ensuring that there is a collective decision-

making process involving both commissioner and service provider, or a neutral referral party/mechanism. This provides the opportunity for all parties to make and agree decisions that are in the best interests of the service recipient and avoid so-called “cherry picking” of the cohort;

- unwanted ‘creaming’ of easier to help cases can also be mitigated by collective scrutiny of who is receiving the intervention, or by avoiding metrics that encourage providers to favour some recipients over others;
- perverse incentives that arise due to specific time-related metrics or binary outcomes can be mitigated by allowing exceptions – or changing the metric to one that rewards success at small, regular intervals, or offers bonus payments rather than according to ‘cliff-edge’ achievement at a single defined point after several months.

## Involving providers and investors

As in many aspects of outcome based commissioning, it is important that the decisions made regarding the outcomes framework are tested with potential service providers and, if involved, social investors.

**Indeed it is arguable that the acceptability of the proposed outcomes framework to all parties is the single most important consideration in terms of finalising a SIB contract.**

If the commissioner is prepared to pay for the achievement of an outcome, and providers and investors are comfortable to have success measured by the achievement of the same outcome, a successful contract is likely. Conversely, if one or more parties are not happy with the proposed outcome measure or the performance level required to achieve payment, it will be very difficult to conclude a contract successfully.

Commissioners may wish to consult other parties while testing the feasibility of the proposed contract, and should certainly do so during the development stage and prior to the conclusion of any competitive process. This will help ensure the acceptability of the proposed framework to all parties, and also enable the commissioner to benefit from providers’ and investors’ previous experience of similar contracts, and the outcomes measures and metrics used within them, where appropriate.

There may then be further negotiation or dialogue with providers and/or investors during implementation, depending on the procurement process used. One of the key considerations in choosing the appropriate procurement procedure may be whether the commissioner believes that there is likely to

be a difference of view between providers and investors on the outcome measures and metrics proposed. If this is the case, it is advisable to select a procedure – such as the restricted procedure with competitive dialogue – that allows tenderers for the contract to suggest or request changes to the outcome metrics proposed.

Where contract development is being led by a provider, it becomes even more important that there is consensus around the outcomes

framework, and the relevant commissioner(s) are comfortable with the measures and metrics proposed. Providers may be more heavily involved in the development of the outcomes framework in these circumstances, but need to be aware that commissioners, as those ultimately paying for outcomes, may want and expect considerable involvement, and often the final say, in the agreement of the measures and metrics used.

## Evaluation

In the majority of SIBs there will be some form of external evaluation of the project, to determine whether the service delivered value and impact compared to a counterfactual case, and to prove that the service actually had the expected level of attribution to the outcomes achieved. In cases where the payment mechanism does not depend on an independent impact evaluation, the evaluation will take a retrospective view of the project after outcomes have been measured and payments made, but it will answer to the points around whether relative value was delivered

and which aspects of the service contributed to the outcomes. Its primary purpose is to create learning for the future. In the UK, payment has rarely been contingent on independently evaluated results.

Evaluation also depends on defining measures and collecting data, so it should be part of the considerations around designing the datasets to be collected and reported.

See the GO Lab guide to Evaluation  
available at [golab.bsg.ox.ac.uk](http://golab.bsg.ox.ac.uk)

# TOOLKIT: SETTING AND MEASURING OUTCOMES

Feasibility  
pp 21–22



Development  
pp 23–26



Implementation  
pp 27–28



Operation  
p 29



Evaluation  
p 30

# Feasibility

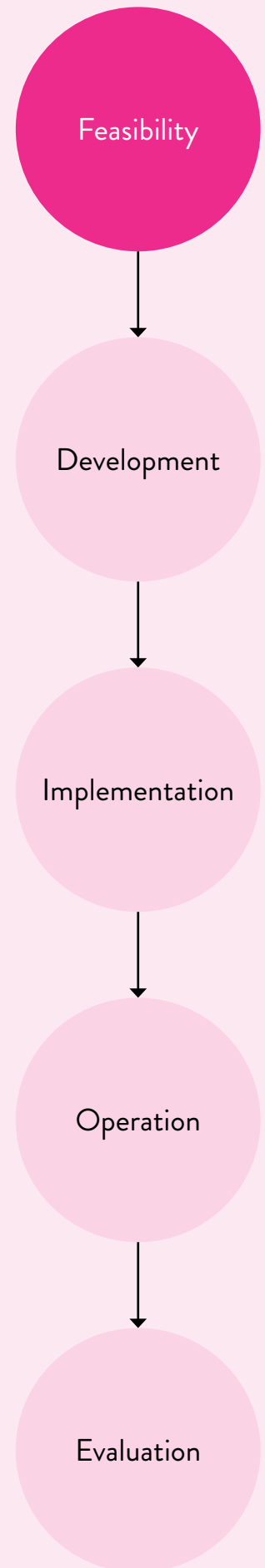
## Setting and measuring outcomes: summary of key actions

	Actions	Considerations
Feasibility	Define the overarching outcomes or objectives or the service	Is there clarity on the overall intent of the service?
	Confirm availability of clear and measurable outcomes	Are there clear outcome measures available which align to policy and financial objectives? Can they be measured at acceptable effort and cost?
	Assess financial and social benefits of outcomes being achieved	What is the current cost of outcomes not being achieved? What will be the value of their achievement?
Development	Define outcomes framework setting out outcome measures, metrics and specific payment triggers where appropriate	What is the best way to measure outcomes and does it require investment in new processes and systems? Does the framework create perverse incentives and how should these be mitigated or avoided?
	Model and define payment structure linked to achievement of outcomes	Is individual measurement against a tariff or cohort-level measurement more appropriate? How much payment should be attached to the achievement of each outcome?
Implementation	Define each outcome and metric in terms that can be included in a specification of requirement and contract	What actions are required and by whom to validate achievement of an outcome or specific metric in terms acceptable to all contract parties?
	Select a procurement procedure that allows for refinement of outcomes, metrics and payments	Which procurement procedure will best allow scope (if needed) for providers or investors to suggest changes to the outcome metrics or the way payments are assigned to metrics?
Operation	Review the measurement of outcomes in the light of contract performance	Does the operation of the contract indicate any need to change the outcome metrics or the way they are evaluated?
Evaluation	Evaluate whether the outcomes have been achieved and whether key assumptions have been validated in practice	Did the contract achieve the level of impact expected and assumed in the outcomes framework and payment structure?

A major consideration in establishing whether it is feasible to put in place an outcomes-based contract, is whether it is possible to specify the social problem that needs to be addressed in terms of clear and measurable outcomes. The improvement of these outcomes for the cohort will also be a key determinant of the financial and social value to be derived from the contract.

The key objectives during the feasibility stage are therefore to:

- confirm that outcomes can be defined that align to the social and policy objectives of the project and assess in general terms whether they are measurable
- estimate the current costs that the cohort is incurring through failure to achieve these outcomes. This is a key component in the overall financial and value case for the contract, along with the impact that the intervention is expected to achieve.



## 1 Confirm the availability and measurability of outcomes

At feasibility stage the objective is to establish the feasibility of an outcomes-based approach by identifying outcomes that address the social problem that are clear and can be measured. These can then be used to engage internal stakeholders and co-commissioners, where appropriate.

### Key actions

- Identify the key outcomes that will drive the contract and payments made
- Confirm that mechanisms exist through which outcomes could be measured

### Key deliverables

- A statement of the outcomes selected and how they might be measured

### Key considerations

- Are outcomes available that are well aligned with the social problem commissioners are aiming to address?
- Which outcomes have the best alignment with project objectives if several are available?
- Which outcomes generate the greatest financial benefit to the lead and other commissioners?
- Are there examples from previous contracts or other best practice on which to draw?

## 2 Estimate the costs incurred by failure to achieve the outcomes

At feasibility stage commissioners should aim to establish the base financial case for the contract and its viability. Part of this is establishing what costs would be incurred by the cohort (the counter-factual) if outcomes were not to be changed by the contracted intervention. There will be high correlation between these adverse outcomes and the main outcomes identified as objectives of the contract as above.

### Key actions

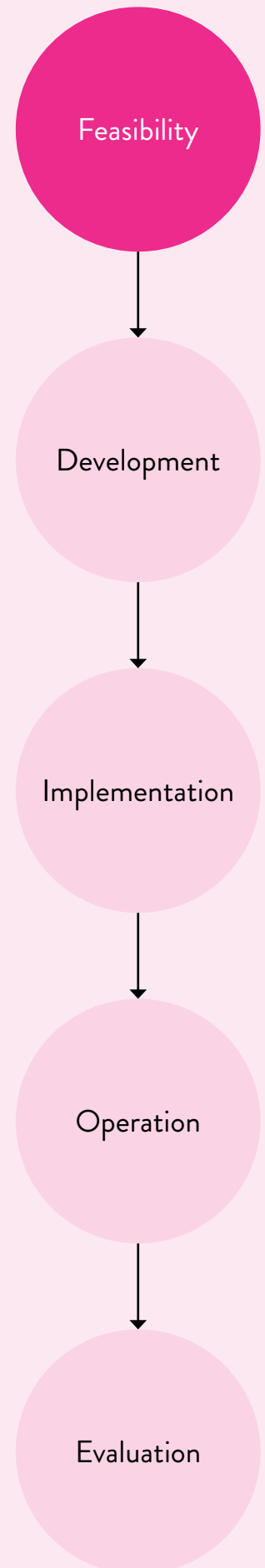
- Estimate the current costs of adverse outcomes to be improved by the intervention

### Key deliverables

- Cost estimates with stated assumptions to feed into financial case

### Key considerations

- What is the best source of information on current costs including local data (e.g. on costs of care placements) or national sources (e.g. New Economy database, PSSRU Unit Costs of Health and Social Care)?
- What assumptions should be made about adverse outcomes as to duration (e.g. length of time workless) frequency (e.g. number of exclusions from school) or severity (e.g. of proven offences committed)?
- Should additional adverse outcomes be costed that are not key outcomes for the contract?



# Development

During the development stage the commissioner should build on the initial statement of outcomes identified during feasibility to develop a complete outcomes framework setting out the outcomes to be achieved by the contract, how they will be measured, and what metrics should be applied to each measure to determine success and trigger payment. The commissioner will also need to develop a payment structure which sets out the payment terms and estimates the level of payment to be attached to each outcome and metric.

The key objectives during the development stage are to:

- define in detail the outcomes to be achieved and how they can best be measured
- specify metrics and triggers for each outcome measure that will drive payment under the contract
- develop a payment structure that sets out the level of payment to be made if each outcome metric is achieved.

## 1 Define the outcomes and how they can best be measured

During the development stage commissioners should consider in detail the outcome measures they wish to use and the implications of using them as regards both how data is collected and reported, and other issues such as whether the outcomes and measures create perverse incentives that need to be mitigated.

### Key actions

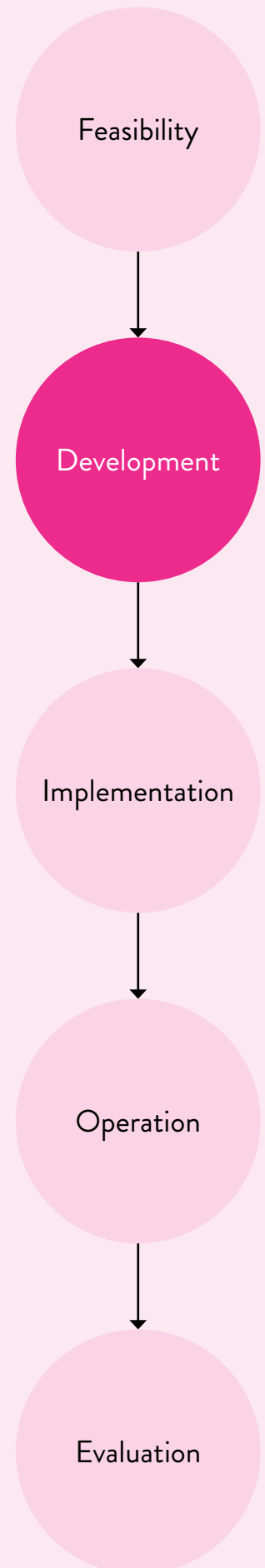
- Refine the outcomes identified into a detailed framework of outcomes and measures that will form the basis of the contract
- Consider the implications of each measure and actions needed to ensure it can be used effectively

### Key deliverables

- Defined outcomes and measures including primary and secondary outcomes, associated measures, and data collection and management requirements

### Key considerations

- How will the data required for each outcome and measure be collected and who will be responsible for its collection?
- Are existing systems and processes for data collection available? If not, what will be the cost of such systems?
- What is the primary outcome that is the most important driver of the contract?
- What are the secondary outcomes and are there any outcomes that will be measured but not linked to payment under the contract?
- What potential perverse incentives are created by each measure and do they need to be mitigated if they cannot be designed out?



## 2 Specify metrics and triggers that will drive payment under the contract

Each outcome measure will need specific metrics and triggers assigned to it. There may be a single metric or several triggers relating to the time over which the outcome is sustained, or the percentage change achieved.

### Key actions

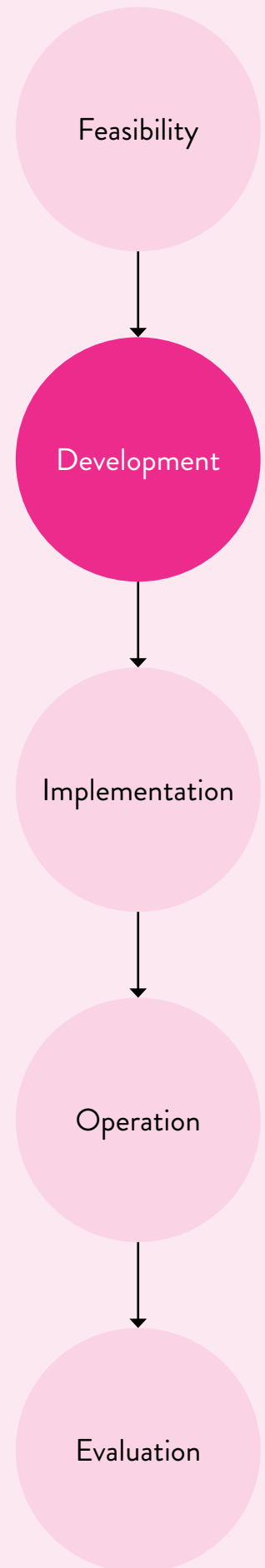
- Identify metrics for each main and secondary outcome that the commissioner intends to link to payment under the contract

### Key deliverables

- Clear metrics and triggers that feed into the overall outcomes framework

### Key considerations

- Should metrics relate to individuals or impact achieved across the cohort as a whole?
- Are interim or progress metrics needed to check and reward progress towards the primary outcomes?
- Should there be metrics that reward sustainment of the outcome and if so, for how long?
- Do the metrics themselves create perverse incentives that need to be mitigated if they cannot be designed out?



### 3 Develop a payment structure that sets out the payments to be made

The payment structure needs to achieve the right balance of risk and reward for both the commissioner and for providers and investors, and also achieve a sensible balance between the primary and other outcomes that the contract is aiming to achieve.

#### Key actions

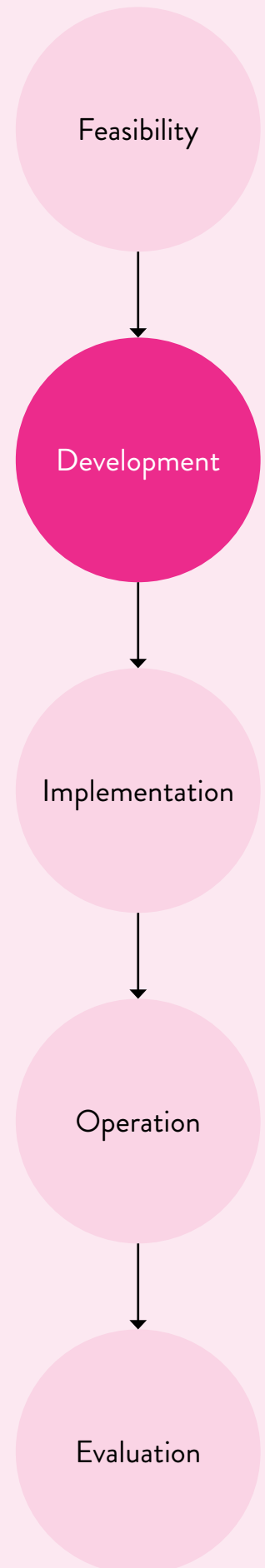
- Develop a financial case that will enable the commissioner to test the effect of different payment levels for each metric on the total payments made and their timing
- Use the model to arrive at a structure that enables the provider and/or investors to cover intervention and other costs, while being acceptable to the commissioners making outcome payments

#### Key deliverables

- A proposed payment structure that can be used as the basis of consultation with the market and feed into procurement

#### Key considerations

- What level of performance can reasonably be expected from the intervention?
- What are the estimated costs of the service, including the intervention, returns to investors and other costs that need to be covered by total payments?
- How much are commissioner(s) willing and able to pay for outcomes, and does this vary according to each outcome and metric?
- How should payments vary across each metric, according to its importance and commissioner preference?
- If a grant to support commissioner payment is available, where is it best applied and what will be the effect of different levels of grant being awarded?



## 4 Engaging and consulting providers and investors

The outcomes framework and payment structure should be designed with involvement from providers and investors who will be able to draw on previous experience of outcome measures and metrics for different types of contract. They will also have views on whether the proposed payment structure achieves a good balance of risk, and issues such potential perverse incentives. Where a provider and/or investors is not directly involved in the co-design of the contract, commissioners should consult a sample of providers and investors to get feedback on their proposals.

### Key actions

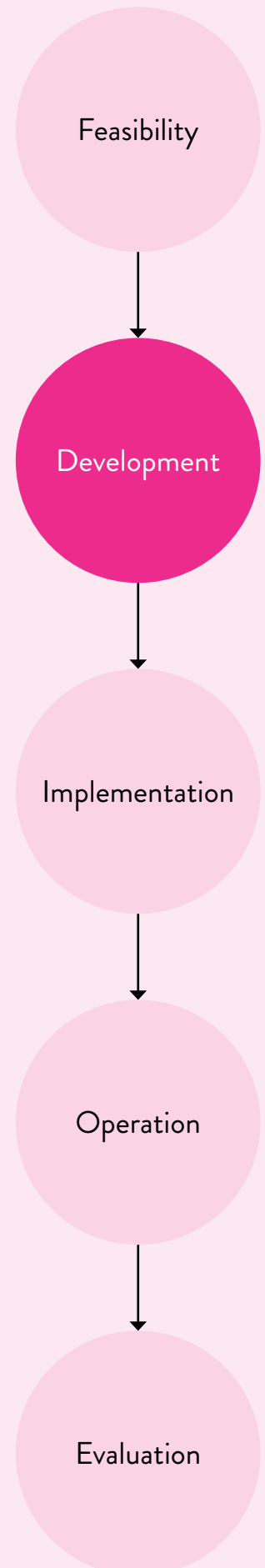
- If not co-designing the contract, document the proposed outcomes framework, metrics and payment structure in a way that can be shared with providers and investors
- Seek feedback from providers and investors either as a group or in one to one discussions
- Revise and refine the metrics and payment structure in the light of feedback received

### Key deliverables

- A revised outcomes framework and payment structure that will feed into procurement

### Key considerations

- How many providers and investors to consult – we recommend 3–4 of each
- Who best to involve – the primary consideration should be those who have previous experience of similar contracts and especially outcomes-based contracts addressing the same or very similar social problems
- How to reconcile conflicting feedback – investors and providers will not always agree, especially about the value of individual outcome payments and the amount of risk that is acceptable to them, so commissioners ultimately need to make their own judgement on whether and to what extent to amend their proposals



# Implementation

In the implementation stage the commissioner will be procuring the contract provider and investor if appropriate. The outcomes framework and payment structure (and associated payment tariff if relevant) will be important elements in the specification of requirement and tender documentation, and may need further detail added – especially regarding how outcomes will be validated and deemed to have been achieved for contractual purposes. There may also need to be scope for investors and providers to negotiate changes to the outcome metrics and payment levels.

The key objectives during the implementation stage are to:

- select and implement a procurement procedure that gives appropriate scope to providers and investors to negotiate changes to the outcomes framework and payment structure
- define outcomes, metrics and validation mechanisms in the specification and draft contract, and ensure those responding to procurement understand and agree the outcomes and metrics proposed, and how their performance will be measured against them.

## 1 Selecting the appropriate procurement procedure

Where the provider and investor are being sought through open competition, commissioners should be careful to choose a procurement procedure that allows providers and investors to influence and, where necessary, change the outcome metrics and triggers, or the values attached to them, during the competitive phase. If a non-competitive process is being used (e.g. Voluntary Ex Ante Transparency notice (VEAT), commissioners should negotiate the framework where possible prior to the award. See the GO Lab Guide to Procurement for advice and guidance.

### Key actions

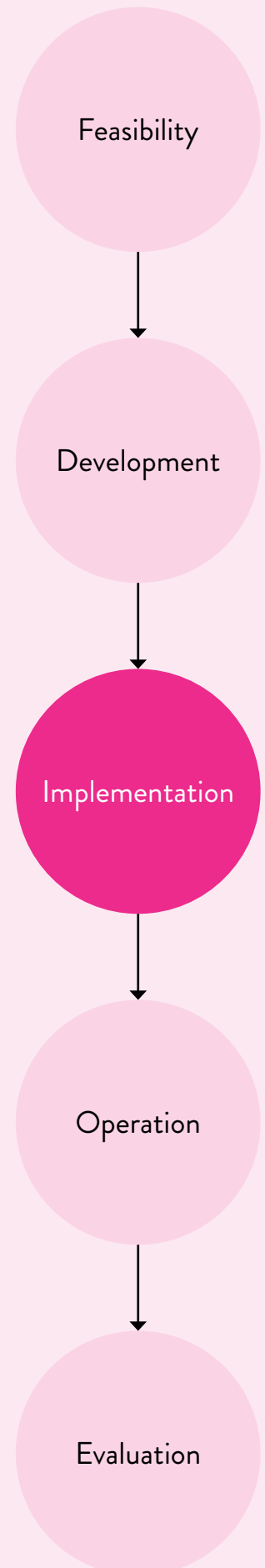
- Consider whether flexibility to negotiate outcomes and metrics is needed during the procurement process
- If it is needed, select an appropriate procurement procedure – note that the choice of procedure has to be justified to those bidding for the contract

### Key deliverables

- Analysis of procurement options with a clear decision on the procedure that has been selected and why
- If time is tight, use of the accelerated version of either procedure may be appropriate – but should be balanced against the need to allow enough time for tendering and any discussions or negotiations to be properly concluded

### Key considerations

- If flexibility is required within an open competition the main options are the restricted procedure with competitive dialogue; and the negotiated procedure



## 2 Defining outcomes in the specification

The specification of outcomes in legal and contractual terms, requires more detailed definition and explication than is usual in an outcomes framework. Commissioners should seek advice from the legal and procurement colleagues on this, and draw on their experience of developing contract documentation

### Key actions

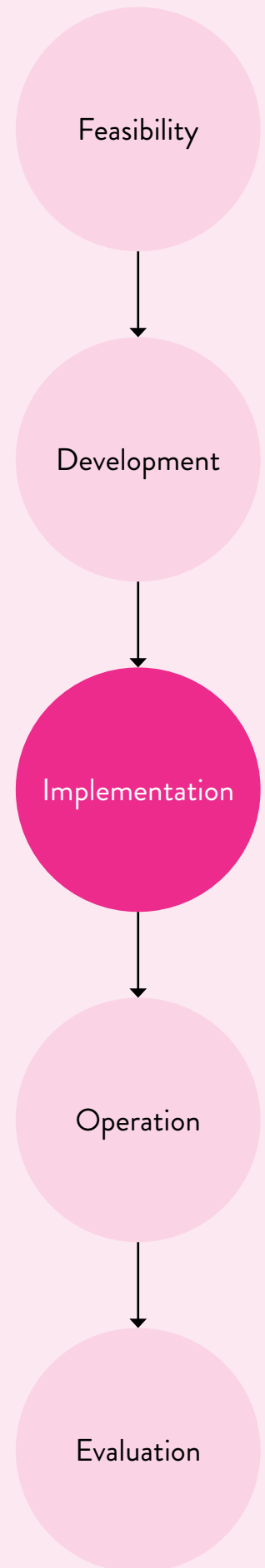
- Define the outcomes and associated metrics in clear and unambiguous terms

### Key deliverables

- Defined outcomes in specification and draft contract

### Key considerations

- Specify the basis on which achievement of an outcome will be validated – e.g. school attendance must be signed off by a head teacher
- Ensure there is no room for doubt or ambiguity between parties to the contract as to whether an outcome has been achieved



# Operation

In the operational stage the commissioner will be actively managing the performance of the contract and responding to the evidence for the achievement of outcomes being generated by the provider.

From the standpoint of the outcomes framework itself, the main action that may be required during the operational stage is to consider changes to the outcomes measures or metrics if it proves more difficult than expected to collect data, or unexpected exogenous factors invalidate or undermine a chosen outcome measure.

The key objective during the operation stage is to:

- review the outcome measures and metrics in use and consider revision if all parties to the contract agree that internal or external changes over the contract period make it necessary.

1

## Reviewing the outcomes framework and the measures and metrics in use

While there will be many contracts where no changes are needed, commissioners, providers and investors should keep the outcomes framework under review through the contract governance structures

### Key actions

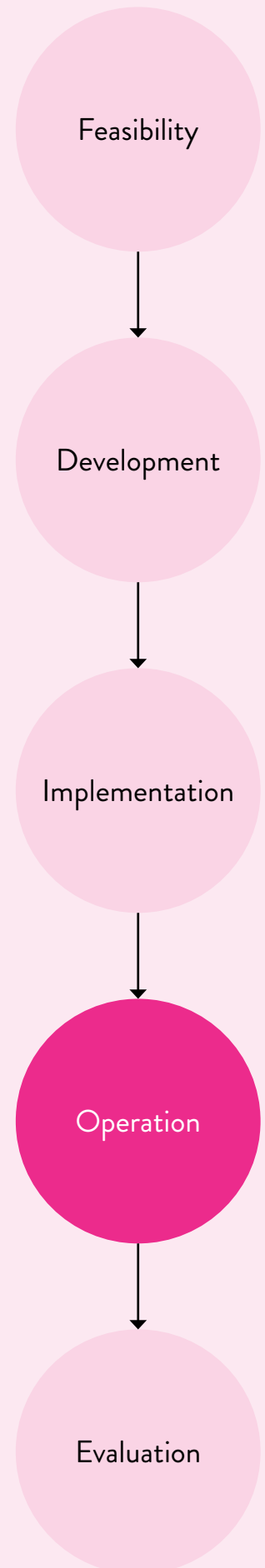
- Review the applicability of the outcomes framework at intervals or by exception if an external event demands it

### Key deliverables

- Change control documents where needed to amend outcomes framework

### Key considerations

- Have external policy conditions changed in a way that undermines an outcome or metric?
- Have there been unexpected drawbacks to the use of some measures and metrics?
- Have external factors impacted a measure or metric to an extent that means either the measure itself or the level of success demanded by a payment metric need to be changed?



# Evaluation

In developing and implementing their strategy for evaluation, commissioners should build in a robust evaluation of the extent to which the outcomes selected have been the right ones to use (part of the process evaluation), and whether the degree of change in outcomes achieved by the intervention was in line with, exceeded or lower than expectations (part of the impact evaluation).

Where the measurement of and payment for outcomes has been tariff-based, it will be particularly important that the evaluation uses a

robust counter-factual to assess the soundness of the assumptions made during the design of the payment structure, especially as regards deadweight.

The key objective during the evaluation stage is to:

- evaluate the appropriateness of the outcomes framework and the robustness of the assumptions made in setting the outcome metrics and payment structure and tariff.

## 1 Evaluation of the outcomes framework and payment structure/ tariff

### Key actions

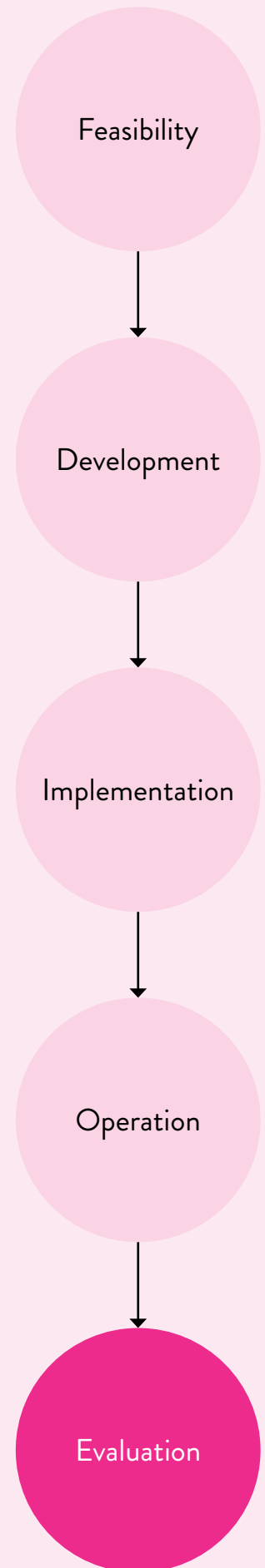
- Develop an evaluation strategy that ensures lessons are learned from the choice of outcome measures and metrics and the calculation of the payments tariff

### Key deliverables

- Commentary on outcomes framework and tariff assumptions in reports on process and impact evaluations

### Key considerations

- Did the outcomes framework provide the right incentives?
- Were there unexpected drawbacks (or benefits) from the choice of outcome measures and metrics?
- Was the impact of the intervention as predicted when the outcomes tariff was set?





**GOVERNMENT  
OUTCOMES  
LAB**

---

**To get in touch with us please email  
[golab@bsg.ox.ac.uk](mailto:golab@bsg.ox.ac.uk) and follow us  
on Twitter for news and updates  
[@ukgolab](https://twitter.com/ukgolab)**

1st Edition

Date of Publication: November 2017