BLAVATNIK SCHOOL OF GOVERNMENT

UNIVERSITY OF OXFORD

GOVERNMENT OUTCOMES LAB

Research

Advise

Connect

Setting and Measuring Outcomes

26th April 2018

@ukgolab
#SIBsOutcomes

About us

Joint partnership between UK Government & Oxford University

Established in 2016

Based at the Blavatnik School of Government, in Oxford

Centre of academic research and practice with a mission to improve the provision of public services to tackle complex social issues, with a focus on outcome based models

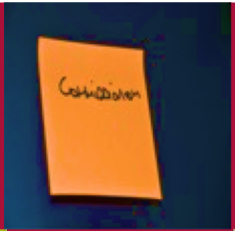
Support available from GO Lab

Knowledge Hub
golab.bsg.ox.ac.uk

Communities of practice
(peer learning)

How to guides

Commissioners' journey tool



Webinars

Advice surgeries

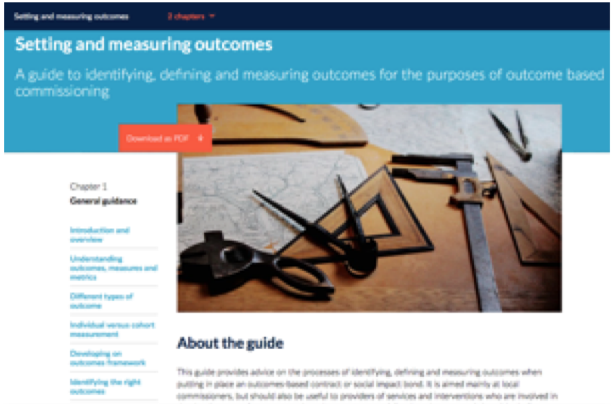
Events & workshops

SIB projects database

Executive education

Fellows of Practice

Some helpful resources



How to guides:
golab.bsg.ox.ac.uk/technical-guides

Book online:
golab.bsg.ox.ac.uk/advice-surgeries

Innovation In Outcomes-based Commissioning West Midlands Regional Conference

22 May, Birmingham

golab.bsg.ox.ac.uk/events

Today's speakers

Dr Clare FitzGerald, Research Fellow, GO Lab

Neil Stanworth, GO Lab Fellow of Practice & Director, ATQ Consultants

Andreea Anastasiu, Policy & Engagement Officer, GO Lab

Research

Advise

Connect

Our audience this morning (Zoom poll)

@ukgolab
#SIBsOutcomes

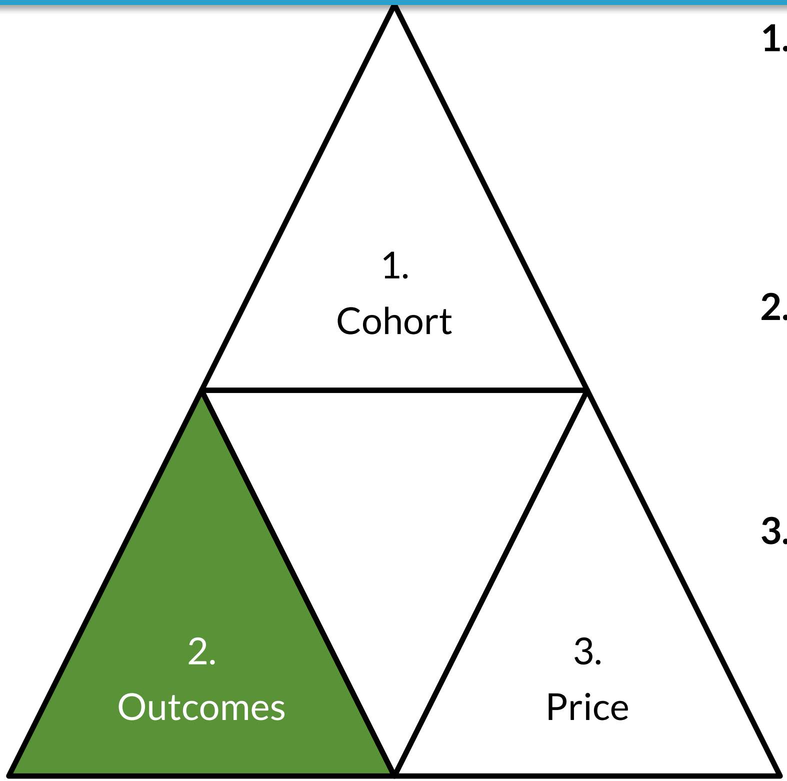
Your main questions were...

- What is the process of developing an outcomes framework?
- How do you get the right balance between stretch and achievability in setting outcome metrics?
- What is the difference between cohort level and individual measurement?
- How do you avoid perverse incentives and cherry picking?
- How can the various parties involved in a SIB negotiate realistic and measurable outcomes?

Session overview

- Understanding outcomes: some key definitions
- Identifying the right outcomes
- Measuring outcomes
- Cohort & individual measurement
- Setting outcome metrics & triggers
- Gaming & perverse incentives

Designing a robust framework



1. **Tightly defined eligible cohort**
   - Clear, objective criteria
   - Understanding of how far participants are from the desired outcomes
   - Independent referral / identification mechanism

2. **Alignment between payable outcomes and policy objectives**
   - Logical link between activity, outputs and outcomes
   - Adequate period of time for tracking
   - A way to tell if the effect has 'stuck'

3. **Accurate price-setting of outcomes**
   - Robust estimate of likely level of benefit vs what would happen anyway ("deadweight")
   - A way to get confidence that any outcomes are caused by the intervention ("attribution")
   - Suitably long outcome tracking-period

Whilst it is not practical for these aspects to be perfect,
commissioners should focus on them to avoid perverse incentives

Understanding outcomes: key definitions
Your main questions were…

- What is the difference between outcome measures and metrics?
- What are the different types of outcomes and when is it appropriate to use them?
- Hard & soft outcomes. Payment amounts in terms of soft and hard outcomes
- What are proxy outcomes and when should these be used? What are the risks in using proxy measures?

Outcomes, measures, metrics

Outcome	What changes for an individual (or other defined unit, such as a family) as the result of a service or intervention (e.g. improved learning outcomes)
Outcome measure *(also termed an indicator)*	The specific way the commissioner chooses to determine whether that outcome has been achieved (e.g. a test score)
Outcome metric *(also termed triggers or targets)*	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 Payment by Results (PbR) contract or Social Impact Bond (SIB), these metrics will usually determine whether a payment is made to the provider or investor.

Outcomes, measures, metrics

Example: Youth unemployment reduction

Outcome	The young person is in employment
Outcome measure *(also termed an indicator)*	Confirmation from the employer that the person is employed by them
Outcome metric *(also termed triggers or targets)*	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

Outcomes, measures, metrics

Example: Emotional wellbeing

Outcome	Improved emotional wellbeing of young people
Outcome measure *(also termed an indicator)*	An identifiable improvement in young people's resilience and ability to deal with challenges using the Strengths and Difficulties Questionnaire (SDQ)
Outcome metric *(also termed triggers or targets)*	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

Outcomes, measures, metrics

Example: Managing long term conditions

Outcome	**People are able to manage their long term condition without hospital treatment**
Outcome measure *(also termed an indicator)*	A reduction in the number of hospital admissions by people receiving support in relation to the specified condition(s)
Outcome metric *(also termed triggers or targets)*	A specific reduction (e.g. 1 per person) in the average number of planned or unplanned admissions (or both) across the cohort of those receiving support, compared to the average prior to intervention or The average admissions by a comparison group

Hard & soft outcome measures

Type of outcome	Definition	Advantages	Disadvantages
Hard outcome	Any outcome that can be measured objectively	▪ 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
Soft outcome	Outcome that requires subjective assessment of its achievement	▪ Useful when a hard outcome is not available ▪ Can be 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

Hard & soft outcome measures

Example: managing long terms conditions

The Ways to Wellness Social Impact Bond (SIB), commissioned by Newcastle Gateshead CCG, measures improvements in the management of long term conditions achieved through social prescription.

It uses a combination of:

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

Proxy outcomes & measures

Proxy outcome measures:

- An indirect measure of the desired outcome which is itself strongly correlated to that outcome, used when direct measures of the outcome are unavailable or cannot be measured.
- **BUT 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 outcomes-based contract*)

Some examples:

- In the health example on the previous slide, reduced hospital admission is itself a proxy measure of the true outcome, which is improved self-management of conditions. However reduced attendance might be due to longer waiting times, or other conditions having higher priority, rather than positive action by the person receiving support
- Being 'off benefit' has been widely used as a proxy for 'in employment', but an individual may no longer be claiming benefits without finding work
- Conviction for an offence is used as a proxy measure for reoffending, but many offences go undetected so it is likely to underestimate both the scale and potentially severity of offences committed.

Lead/progression measures examples

- Improved school attendance and/or behaviour
- Engagement in part time or voluntary work
- Family attendance at a parenting support programme.

May lead to

- Improved attainment and reduced risk of exclusion
- Full time employment
- Reduced risk of a child becoming 'in need' or on a Child Protection Plan.

BLAVATNIK SCHOOL OF GOVERNMENT

UNIVERSITY OF OXFORD

GOVERNMENT OUTCOMES LAB

Research

Advise

Connect

Live Q&A

@ukgolab
#SIBsOutcomes

Identifying the right outcomes
Your main questions were…

- What is the process of developing an outcomes framework ? Where do you start from?
- How do you identify the right outcomes to use? Strategic v operational outcomes – which are best?
- How are outcome measures co-created? How do you include service users in setting outcomes?
- Can you revise your outcome framework after delivery has started? How do you leave space for flexibility in setting the outcomes?

Developing an outcomes framework

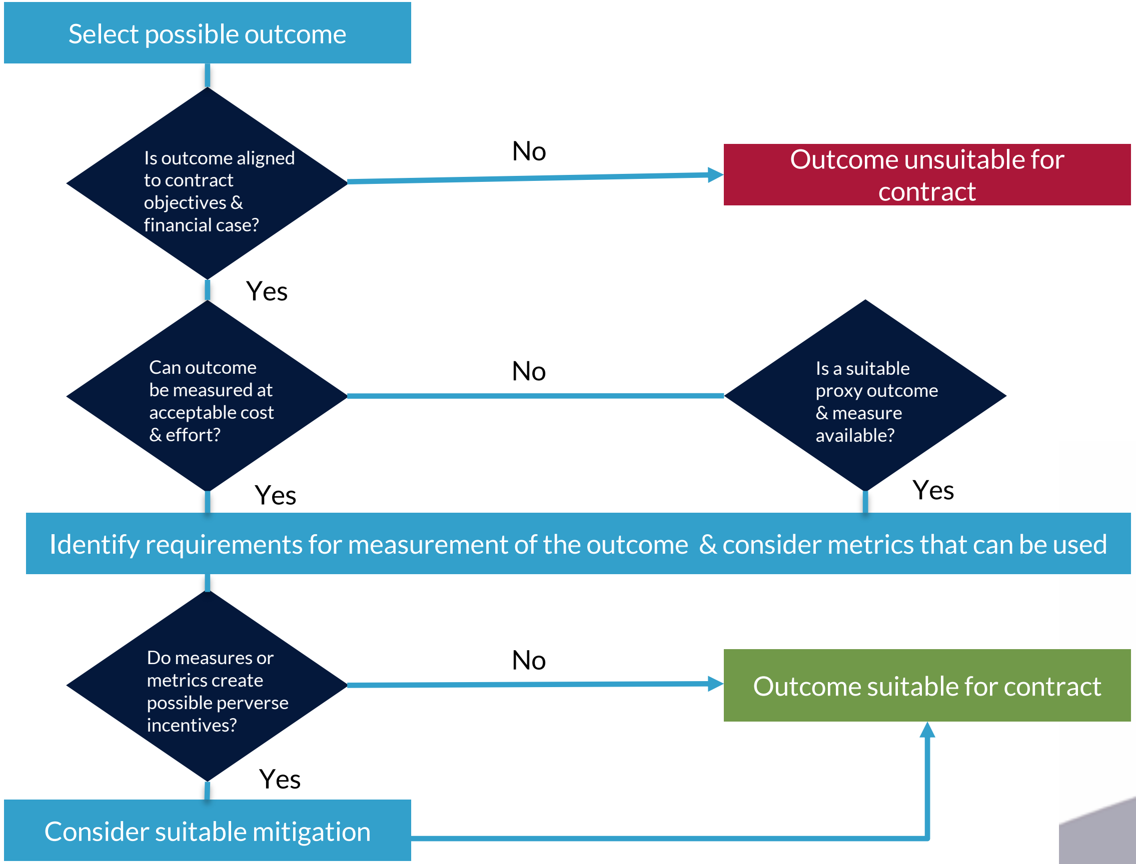
An outcomes framework should include:

- ✓ 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.

- ✓ Iterative and progressive process
- ✓ Negotiation & consensus-building around the outcomes framework among the contracting parties
- ✓ Review and revise (if necessary)

Identifying outcomes

A simplified decision-making tree



Measuring outcomes
Your main questions were…

- How do you measure outcomes? What are the options? How do you create a measurement system which is robust but not so top heavy that is not fit for purpose?
- How can the various parties involved in a SIB negotiate realistic and measurable outcomes?
- How does the payment mechanism work if the savings have been achieved in a different organisation?
- How can outcomes be measured in complex contexts (e.g. where outcomes cannot be attributed to one single project)?

Measuring outcomes

Some general considerations

- ✓ whether the outcome is best measured using an individual or collective (cross-cohort) approach
- ✓ 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

Data collection

Data type	Pros	Cons
Administrative data	• Highly accurate • Low cost	• May not exist • May not cover population of interest • May not directly address question of interest
Primary data	• Addresses question of interest	• High cost • Possibility of bias

✓ Find out whether the required data is already collected for other purposes

✓ Do not to make assumptions about the availability of data from other parties or the ability of those parties to collect data on your behalf.

Cohort v individual measurement
Your main questions were…

- What is the difference between cohort level & individual measurement?
- What are the respective benefits and limitations of the two approaches?
- Examples

Cohort v individual measurement

Cohort outcome measurement	Individual 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. ▪ 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. ▪ does not normally require a separate calculation of deadweight	▪ 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). ▪ does not involve a comparison group or other baseline against which success can be measured for the purposes of payment. ▪ Tariff or rate card specifies what will be paid per individual per outcome ▪ Requires good evidence of the likely level of deadweight

Cohort v individual measurement

Examples

HMP Peterborough SIB

- 7.5% reduction in reoffending across all SIB cohorts, against a national comparison group
- Find out more about the methodology used to measure the impact of the SIB pilot [here](#).

Innovation Fund DWP Rate Card

DWP pays for one or more outcomes per participant which can be linked to improved employability. A definitive list of outcomes and maximum prices DWP was willing to pay for Round 2 is as follows:

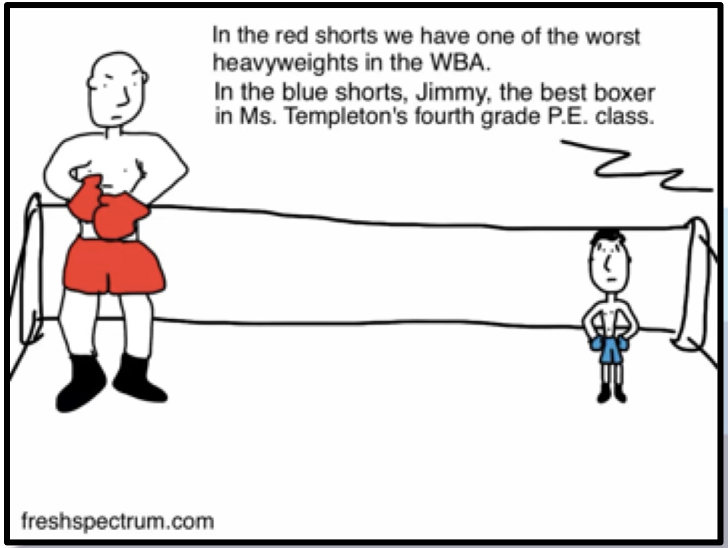
NATURE OF OUTCOME	MAXIMUM PRICE OF OUTCOME
Improved attitude towards school	£700
Improved behaviour	£1,300
Improved attendance	£1,400
Entry level qualification	£900
NVQ Level 1 or equivalent	£1,100
NVQ Level 2 or equivalent	£3,300
NVQ Level 3 or equivalent	£5,100
Entry into employment	£3,500
Sustained employment	£2,000

The maximum amount payable per individual is £11,700. The figure is based on 3 years of Annually Managed Expenditure (AME) savings.

Counterfactuals

Counterfactual: A measure of what the outcome would have been for programme participants if they had not participated in the programme

- 'What would have happened in the absence of the intervention?'
- By definition, the counterfactual cannot be observed. Therefore, it must be estimated using a comparison group
- The challenge – identify a treatment group and a comparison group that are statistically identical, on average, in the absence of the programme.



Source: Chris Lysy, freshspectrum.com/what-is-evaluation-anyway

Deadweight

Deadweight: outcomes which would have occurred without a policy, programme, or intervention.

- *'Did this programme make a difference or would changes have occurred anyhow?'*

Source: Chris Lysy, freshspectrum.com/attribution

Rate cards

A rate card is a schedule of payments for specific outcomes a commissioner is willing to make for each beneficiary/ service user that verifiably achieves each outcome

- Majority of UK outcomes-based contracts have used a rate card or tariff linked to individual outcomes
- Cheaper to manage & usually allows payments to be made more quickly
- Requires robust understanding of what have been achieved without the intervention, to ensure that the commissioner doesn't pay for outcomes that would have been achieved anyway

NATURE OF OUTCOME	MAXIMUM PRICE OF OUTCOME
Improved attitude towards school	£700
Improved behaviour	£1,300
Improved attendance	£1,400
Entry level qualification	£900
NVQ Level 1 or equivalent	£1,100
NVQ Level 2 or equivalent	£3,300
NVQ Level 3 or equivalent	£5,100
Entry into employment	£3,500
Sustained employment	£2,000
The maximum amount payable per individual is £11,700. The figure is based on 3 years of Annually Managed Expenditure (AME) savings.	

Source: DWP Innovation Fund (Round 2)

Setting outcome metrics & triggers
Your main questions were…

- How to encourage stretch targets without discouraging providers?
- How do you balance the requirements of a rigours approach with the challenges and limitations on the ground? (pragmatic versus robust)
- How do you get the right balance between stretch and achievability in setting outcome metrics?

Setting outcome metrics and triggers

- Metrics or triggers specify the level of against which success will be judged/ outcome payments will be made
- Ensure metrics align to the social objectives of the contract and the anticipated financial results
- Metrics should take account of existing evidence around the effectiveness of the intervention and the effect of deadweight
- In some cases, the outcome metric will be the simple achievement of a hard outcome
- In most cases, however, commissioners should consider whether to set a number of additional outcome metrics that reward:
  - Progression towards the main outcome
  - The sustainment of the outcome and/or further improvement

Setting outcome metrics and triggers: practical considerations

The evidence about the level of performance that 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.*

Involving providers (and social investors) in the decision-making process

- ✓ *All stakeholders (commissioners, providers and, if relevant, investors) need to trust there is an objective mechanism for measuring outcomes and that the associated triggers are appropriate. 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.*

Impact of uncontrolled external factors (e.g. possible changes in government policy; planned service changes; changes in local practice)

- ✓ *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.*

Gaming & perverse incentives
Your main questions were…

- Examples of perverse incentives
- How do you avoid cherry picking?

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

Some examples:

- Measures and metrics that link to absence of or reduction in referrals to a statutory agency
- A simplistic binary outcome measure, which can easily be ‘failed’ by a high proportion of the cohort – this can lead to ‘parking’ of those who can no longer achieve the outcome
- Setting a metric that is achieved with varying degrees of effort for different members of the cohort – this can lead to creaming (focussing the attention on those easier to help)
- A single time-related milestone at which the person receiving the information is deemed to have achieved a positive outcome – sometimes termed cliff edge

How to avoid creating perverse incentives

- Referrals: ensure that there is a collective decision-making process involving both commissioner and service provider, or a neutral referral party/mechanism.
- Unwanted 'creaming' of easier to help cases: 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: allow exceptions – or change 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.

Setting outcome metrics & triggers

Example: children on the edge of care

Metrics relating to the prevention of entry to care have evolved through experience and negotiation from:

- Cohort-wide measurement of the total number of days in care compared to a baseline (complex and time-consuming to measure)

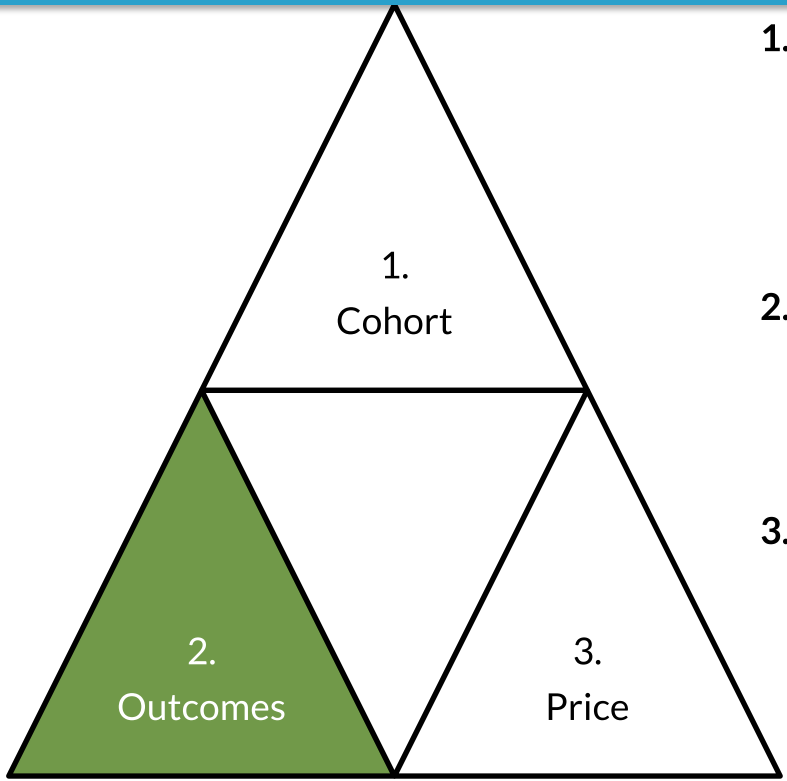
  Through

- Bullet payments paid as an individual tariff at intervals (e.g. 6 months) reflecting the length of time out of care (simpler but prone to distortion or perverse incentives)

  To

- A payment per individual for each care day avoided, totted up and paid at intervals (combining the advantages of a tariff with flexibility and avoidance of distortion)

Designing a robust framework



1. **Tightly defined eligible cohort**
   - Clear, objective criteria
   - Understanding of how far participants are from the desired outcomes
   - Independent referral / identification mechanism

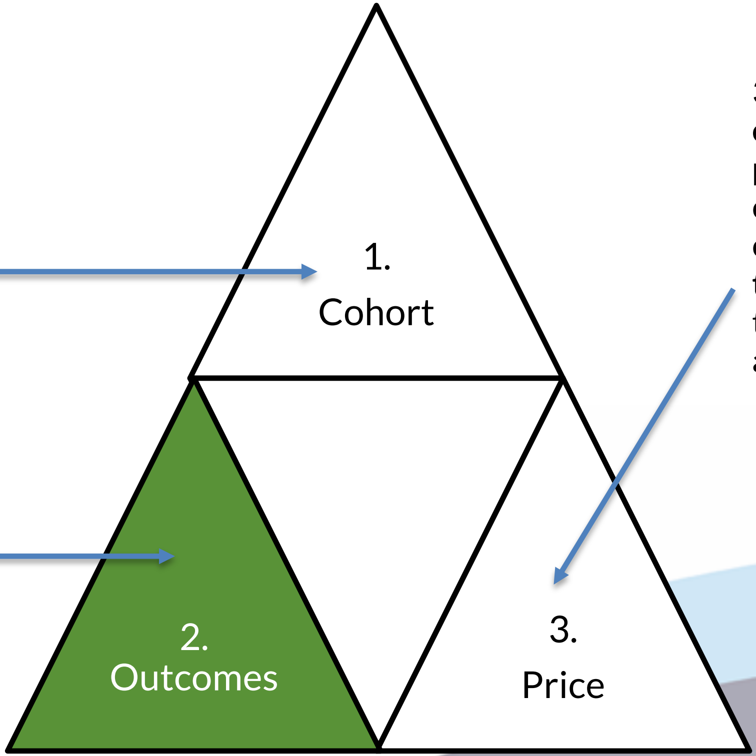
2. **Alignment between payable outcomes and policy objectives**
   - Logical link between activity, outputs and outcomes
   - Adequate period of time for tracking
   - A way to tell if the effect has 'stuck'

3. **Accurate price-setting of outcomes**
   - Robust estimate of likely level of benefit vs what would happen anyway ("deadweight")
   - A way to get confidence that any outcomes are caused by the intervention ("attribution")
   - Suitably long outcome tracking-period

Whilst it is not practical for these aspects to be perfect,
commissioners should focus on them to avoid perverse incentives

What does ‘good’ look like?
Example: HMP Peterborough SIB

1. Eligibility **criteria clear and impartial** (offenders in Peterborough prison sentences of less that 12 months over three cohorts of 1,000) and **referral was independent** (the service was offered to everyone who fit the eligibility criteria, participation was voluntary).



2. Policy **intent to reduce reoffending** and the **project paid on a cohort-based reduction in reconviction**, a valid proxy measure tracked over a suitably long period to be durable.

3. Includes **real-time comparator group in the payment mechanism**: commissioners can say with confidence that the **outcomes they paid for are attributable** to the ONE service, **and additional** to business as usual

Research

Advise

Connect

Live Q&A

@ukgolab
#SIBsOutcomes

Neil's top tips for commissioners

- ✓ Be prepared to iterate and keep iterating - your first thoughts on the best measures and metrics won't be your last
- ✓ Metrics and payments will change when you consult providers and investors – make sure you allow for this, whether during development or procurement
- ✓ Use existing metrics where they are available from other projects - they save time & effort and are more likely to be more acceptable to providers and investors

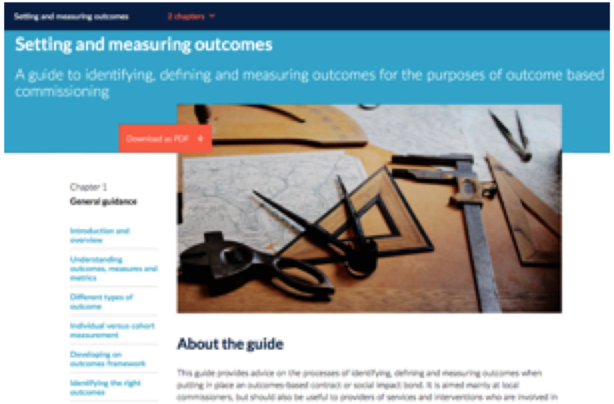
Clare’s top tips for commissioners

- ✓ Think of evaluation as a way to mitigate risk to commissioners, not as a tick box exercise.

- ✓ While you may not elect to move forward with an impact evaluation, the thinking behind what constitutes a good one is a helpful framework from which to design a robust value for money case.

- ✓ The technical bit and the relationship bit are equally important. Nailing one isn’t good enough

Some helpful resources



How to guides:
golab.bsg.ox.ac.uk/technical-guides

Book online:
golab.bsg.ox.ac.uk/advice-surgeries

Innovation In Outcomes-based Commissioning West Midlands Regional Conference

22 May, Birmingham

golab.bsg.ox.ac.uk/events

Research

Advise

Connect

Next steps & your feedback

(Zoom poll)

@ukgolab
#SIBsOutcomes

Stay in touch

@ukgolab

http://golab.bsg.ox.ac.uk

golab@bsg.ox.ac.uk

linkedin.com/in/go-lab-395513140/