Research

Advise

Connect

Setting and Measuring Outcomes
Social Justice Commissioning Academy
3rd July 2018

@ukgolab
#SIBsOutcomes

Introducing us

**Nigel Ball, GO Lab
Deputy Director**

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

About the GO Lab

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
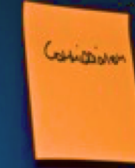
Support available from GO Lab

Knowledge Hub
golab.bsg.ox.ac.uk

Communities of practice (peer learning)

How to guides

Commissioners' journey tool

Webinars

Advice surgeries

Events & workshops
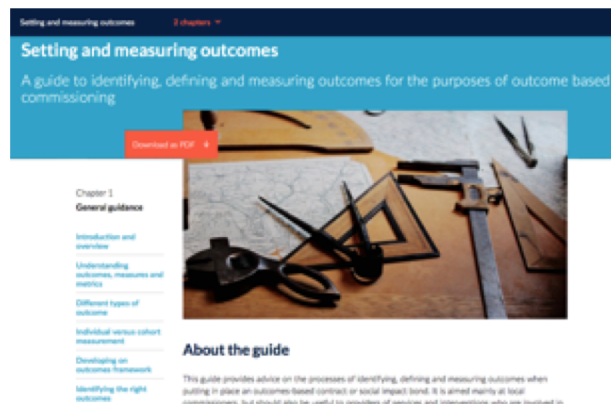
SIB projects database

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

Comparing SIBs and outcomes-based approaches

International Conference

6-7 September, Oxford

golab.bsg.ox.ac.uk/events

Research Advise Connect

Our audience this evening

@ukgolab
#SIBsOutcomes

Why set and measure outcomes?

1. To **manage performance** / learn how to get better

2. To **evaluate** whether something works (but not *why* it works – or doesn't)

3. To provide a means for **payment**

4. Because I've been told to

Key concepts

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)
Measure / Indicator	The specific way the commissioner chooses to determine whether that outcome has been achieved (e.g. a test score)
Metric / Target / Trigger	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).

Outcomes vs Impact

Impact measurement answers that "what would have happened anyway?" question

SSIR - Ten Reasons Not to Measure Impact—and What to Do Instead
https://ssir.org/articles/entry/ten_reasons_not_to_measure_impact_and_what_to_do_instead

Neil Stanworth

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

Key issues and questions

- Hard v soft outcomes

- Binary v continuous measures

- Using proxy and lead indicators

- Cohort v individual measurement

Hard & soft outcome measures

Type of outcome	Definition	Advantages	Disadvantages
Hard	Can be measured objectively	▪ Simpler to measure ▪ No risk of disagreement about achievement	▪ Not always available ▪ May not capture sustained impact ▪ May not reflect what matters to service users
Soft	Requires subjective assessment	▪ Useful when no hard outcome is available ▪ Can be used to test progress made ▪ Measures whether the service meets user expectations	▪ Consistency of measurement can be difficult ▪ Potential for disagreement about achievement

Binary and continuous measures

Type of outcome	Employment	Offending	Child Protection
Binary	In work/not in work	Not convicted/convicted	Looked after/not looked after
Continuous	Length of time in work Quality of job Progression to full time work	Reduced frequency Reduced severity	Length of time not in care Escalation/de-escalation to formal Child Protection

Proxy outcomes & measures

An indirect measure of the desired outcome strongly correlated to that outcome, used when direct measures of the outcome are unavailable or cannot be measured.

Proxy outcomes & measures

Proxy measure	True outcome	Risks
▪ Reduced hospital admissions	▪ Improved health	• Reduced attendance might be due to other factors
▪ Off benefit	▪ In employment	• Person may cease claiming benefits without finding work
▪ Reduced reconvictions	▪ Reduced reoffending	• Many offences go undetected

Lead/progression measures

- 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.

Cohort v individual measurement

Cohort outcome measurement	Individual outcome measurement
▪ Works best when the current adverse outcomes vary across the cohort ▪ Usually requires comparison with a group who did not receive the intervention ▪ Does not normally require a separate calculation of deadweight	▪ Works best when the cohort are experience similar adverse outcomes ▪ Does not usually involve a comparison group or other baseline ▪ 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 compiled using Propensity Score Matching

- 10% reduction in individual annual cohorts
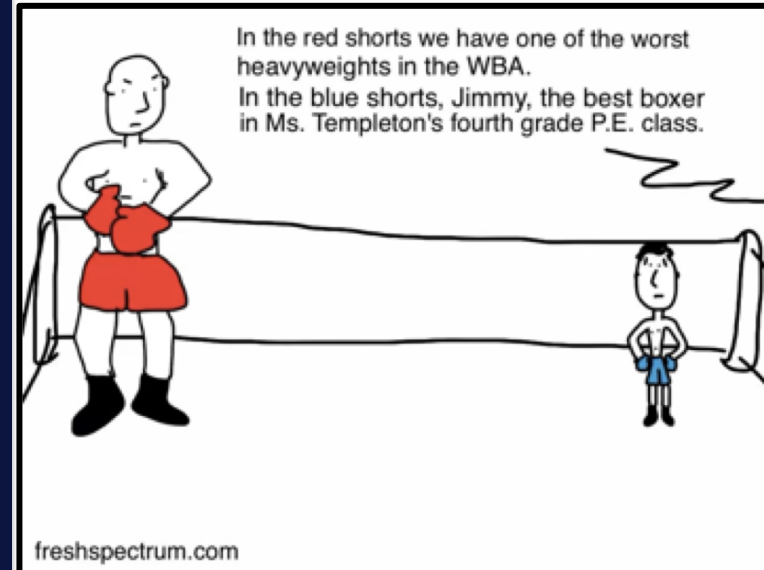
DWP Innovation Fund

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.

Counterfactual

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

- What would have happened in the absence of the intervention

- Cannot be observed, so must be estimated using a comparison group

- The challenge is to identify a treatment group and a comparison group that are statistically identical

In the red shorts we have one of the worst heavyweights in the WBA.
In the blue shorts, Jimmy, the best boxer in Ms. Templeton's fourth grade P.E. class.

freshspectrum.com

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

Deadweight

Deadweight: outcomes which would have occurred without the programme or intervention.

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

- *How many /what proportion of outcomes would have been achieved anyway?*

Source: Chris Lysy, freshspectrum.com/attribution

Data collection options

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	• Directly 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.

How we have learnt to set better outcomes
Example: children on the edge of care

Metrics relating to the prevention of entry to care have evolved from:

- Cohort-wide measurement of the total number of days in care compared to a baseline or comparison group (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)

- A soft measure is often used alongside this e.g. the Family Star

Three top tips for commissioners (and others)

✓ Simple is not always best

✓ Avoid superficially attractive hard and binary metrics that can cause major problems

✓ Be flexible

✓ Measures and metrics will change as you develop your framework and talk to stakeholders

✓ Learn from others' experience

✓ Use existing metrics where they are available – they save time & effort

Research   Advise   Connect

Questions

@ukgolab
#SIBsOutcomes

Stay in touch

@ukgolab

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

golab@bsg.ox.ac.uk

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