Designing Evaluations to Inform Action in New Settings

Calum Davey (and Inception-paper co-authors)
“Use evidence!”

Case-specific
Same problem, many contexts
How, why, and where, as well as what ‘works’

**Aim:** suggest ways to address the issue of learning more from evaluations

**Key idea:** intervention-centred and context-centred thinking

**Experiment:** a semi-context-centred presentation

With a democratic process

But it won’t *work*
**Steps:**

1. Glance at the options
2. Chance to voice a position in favour of either option
3. Close your eyes
4. Raise your hand when the option you prefer is read out
How **best** can we learn from:

- Large-scale, multi-component initiatives to improve the education system in a single country
- Response to Ebola in West Africa for future outbreaks, outbreaks of other diseases, or more generally about how health promotion can be reconciled quickly with cultural norms and expectations?
- What can be learned from the peace-process in Northern Ireland that could be applicable in South Sudan?
- Mobile phone technology to change behaviours, both for future mobile-based interventions but also as a platform for understanding how habits can be changed efficiently?
METHODS

**Co-authors** | **Discipline**
--- | ---
Nancy Cartwright | Philosophy
Edoardo Masset | Economics
Macartan Humphreys | Political Science
Audrey Prost | Anthropology
David Gough | Evidence Synthesis
Sandy Oliver | Evidence Synthesis
Syreen Hassan | Public Health
James Hargreaves | Public Health
Chris Bonell | Public Health
Calum Davey | Public Health

- **Consultations**
- Centre for evaluation members (50)
- DFID Staff (1.5 hrs)
- Intellectual Leadership team (co-authors)

- **Review**
- Expert recommended papers (57)
- Backwards citation (84)
- Forwards citation (ongoing)

- **Writing workshop**
<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fidelity of form</td>
<td>Extent to which form of the intervention (activities, materials, delivery) maintained (or ‘core’?)</td>
<td>Tension between remaining completely faithful to how an intervention was previously delivered (fidelity of form) and making implementation feasible through adaptation</td>
</tr>
<tr>
<td>Feasibility of implementation</td>
<td>Extent of the barriers and facilitators for delivering an intervention in a new setting</td>
<td>Interventions can be described in terms of how they function within a complex system</td>
</tr>
<tr>
<td>Adaptation</td>
<td>Changes to an intervention in a new setting</td>
<td>Interventions can be described in terms of how they function within a complex system</td>
</tr>
<tr>
<td>Fidelity of function</td>
<td>Extent to which the function of an intervention can be maintained in different settings</td>
<td>&gt;3 concepts used for extrapolating causal effects from an evaluation to other places, which differ primarily in the extent to which a particular target context is considered.</td>
</tr>
<tr>
<td>Interventions as events in systems</td>
<td>Interventions as events that interact with systems of people and institutions, no fixed form or function</td>
<td></td>
</tr>
<tr>
<td>Generalisability</td>
<td>Extent to which claims about a sample can be applied to a population</td>
<td></td>
</tr>
<tr>
<td>Transferability,</td>
<td>Extent to which claims made about one setting can be applied in another</td>
<td></td>
</tr>
<tr>
<td>Transportability</td>
<td>(1) Extent to which an intervention can be moved from one place to another; (2) Extent causal inferences about one system can be used to make inferences in another</td>
<td></td>
</tr>
</tbody>
</table>


MECHANISMS & THEORY

CMOs
Elster mechanisms
“Middle-range theory”
MRT, mechanisms, and laws

Is middle-range theory, the solution?
- Can’t handle:
  - Many programme elements
  - Long causal chains
  - Interactions

MRT and programme theory: example

Grand Theory
Structuration
Supply and demand

Mid-range theory
Role conflict
Tendency to cooperate
'entrepreneurs tend not to expand jobs if the available workers have been unemployed for a long period and have lost skills'

Programme theory
Context specific
Derived from more than one MRT

Programmes are constructed, as manifestations of the various theories

---

Context/support factors

Healthcare workers visit schools in rural Kenya
Children in years 1-8 take pills
Worm burden is reduced
Illness is reduced and weight is gained
School attendance increases
Academic attainment increases
Consider the following three theories over binary variables $Y, X_1, X_2$

Theory 1 ('grand theory'):

$$Y = X_1 X_2$$

Theory 2:

$Y = X_1$ in cases in which $X_2 = 1$

Theory 3:

$X_1 = 1$ caused $Y = 1$ for individual $i$ in context $X_2 = 1$

- Theory 1 provides a complete description of a simple process (e.g. children attend school when healthy and school available)
- Theory 2 is implied by Theory 1 -- that is, if Theory 1 is true so is Theory 2, but the converse does not hold (e.g. school attendance equals health status when schools are available)
- Theory 2 provides an incomplete description; here it is a conditional theory,
- Theory 3 is a specific claim about causal relations for an individual (e.g. being healthy caused a child to attend school when there was a school available) Theory 3 claims no generality.

- Within this set of theories Theory 2 is “middle-range”; it is implied by one theory and implies another. In a broader set of theories however Theory 1 might be midlevel, if for example it is implied by some Theory 0.
Maths models of systems

**Behavioural models (economics)**
- Equations are used to describe the complex causal relations
- Combines ‘economic’ structure (deterministic econ theory) with statistics
- May imposes a solution by assuming maximisation of utility or profits or social welfare → make predictions

Both combine assumptions about *structure* with data

Both tend to be applied where implementation is unproblematic (e.g. taxes, vaccines)

**BIQQ**
- Promising method that used Bayesian methods to combine qualitative and quantitative data

**Mathematical models**
- Combine theory (e.g. R, R0) with data
- Present current understanding
- Usually some kind of agent or state model
- Rarely include social processes
- Common sort of question: ‘what would happen to HIV incidence if we successfully treated STIs in this context?’


Context matters: CMOs, support factors

“A short absence can do much good.”
“Absence lessens moderate passions and intensifies great ones, as the wind blows out a candle but fans up a fire.”

Voodoo that works

Lipsey: different models of bullying in primary and secondary schools

PolSci: democracies won’t go to war

Diagnosis

‘Support factors’ (e.g. food, comorbidities)

Short of breath, wasted, pale

Persistent cough (signs, symptoms)

Culture

Lipsey MW (2009) The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*

<table>
<thead>
<tr>
<th>Approach</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test theories not interventions</td>
<td>Orienting evaluations to accumulate knowledge that refines theory, rather than testing the effects of interventions.</td>
</tr>
<tr>
<td>Integrated mixed-methods process evaluation</td>
<td>Gathering data on multiple elements: components of the intervention; implementation; mechanisms (mediators) and effect of context (moderators); representativeness of samples; risk factors; features of target place; practitioner experiences.</td>
</tr>
<tr>
<td>Leverage heterogeneity</td>
<td>Conducting multi-site and pragmatic trials to test assumptions in multiple contexts.</td>
</tr>
<tr>
<td></td>
<td>Using case studies to identify conditions where diverse outcomes are observed focusing on context, implementation differences, and “trajectories of change” to predict whether replication or scale up is possible.</td>
</tr>
</tbody>
</table>
Leverage heterogeneity

Causal case studies
- Causal density
- Process tracing, symptoms, side effects
- Small-n

Multi-site trials
Sample for difference
Shadish, Cook, Campbell Bonell: systematic CMOs

Pragmatic?
*efficacy* of delivery
Voucher only
Not of primary interest to policy-makers


Testing theory

Interventions are imperfect tests
Thinking against oneself

Bonell and bullying:
Underpowered analysis
Potentially confounded mediator analysis

Glennerster and JPAL:
The J-PAL framework begins with the question, ‘what is the disaggregated theory behind the program?’, where ‘theory’ is described as ways of simplifying the world to help make and test predictions. The framework then focuses on the local conditions (and whether the theory is likely to apply), the evidence available to support the behaviour-change underpinning the intervention theory, and considers the evidence that the programme can actually be implemented in this setting. With reference to cases from various contexts, the authors argue that superficial contextual differences might not be important when mechanisms travel and that learning about theory will require synthesis methods that interrogate the mechanisms behind the interventions.

Blattman C (2016) Why what works is the wrong question - Evaluating ideas not programs. URL: https://chrisblattman.com/2016/07/19/14411/
INTRO

METHOD

MECHANISMS & THEORY

MARKERS

EVALUATION APPROACHES

ADAPTIVE DESIGN

DISCUSSION

Definitions

Maths. example

Maths. models of systems

Leverage heterogeneity

Testing theory

Outcome theory

DFID examples

Synthesis

Left of the ToC

Transport formula

ROUTE MAP
Outcome theory

Using evaluation to ask: why does the outcome occur at all?
Randomisation = exogenous variation
IV
Burchett, Ghana
Local understanding of problems
Efficacy studies
Stephen Birch: policy makers care about how health comes about
Shift focus from providers and services
To populations and problems


ADAPTIVE DESIGN

The ‘intervention’ is the response to context
Formative?
Democratic?

‘the intervention’ =

Drawing on evidence

Activities

Context

Examples
Women’s groups
Adaptive management
Wang et al. (2005) suggested that a ‘list of attributes that may impact on applicability and transferability can be developed, based on knowledge of the proposed intervention. Then the applicability and transferability of the intervention to the local setting can be rated, and given a score, based on knowledge of the local setting’

Burchett – none of the frameworks work

**Design focus**
- Design strategy for “hybrid prevention programmes” that “build in” adaptation to enhance programme fit (Castro)
- ADAPT-ITT model (Wingood and DiClemente)
  - Eight phases, including formative research
ROUTE MAP

INTRO

METHOD

MECHANISMS & THEORY

MARKERS

EVALUATION APPROACHES

ADAPTIVE DESIGN

DISCUSSION

Definitions

Maths. example

Maths. models of systems

Leverage heterogeneity

Testing theory

Outcome theory

DFID examples

Synthesis

Left of the ToC

Transport formula
When are causal inferences estimable from both the statistical information available and the causal information transferred from the experiments?

Use ‘selection diagrams’, extensions of DAGS; $S$ depicts the things that differ

Requires considerable confidence in one’s understanding of the causal model in both populations in order to specify the key ways in which they differ.

Requires a minimal compatibility between graphs.

Nodes are assumed to be identical in different contexts, which is a challenge for construct validity, and may require that latent variables are added.
Reminder: most things don’t ‘work’
Contexts vary a lot

When things don’t work, are we looking in the right place?

EEF example

Decision / design processes

Contextual information

How did this ToC, this intervention, get applied here in the first place?

Why do we think of the ‘intervention’ as starting after the design is finished?
DISCUSSION

Summary
- Interventions are based on theories, and evaluations can help strengthen them
- Need to embrace heterogeneity
- Don’t be scared of assumptions (=theory)
- Need to think about how to respond to context in evidence-based way

Context-centred delivery
- Not sure if it ‘worked’ (counterfactual = normal delivery)
- Not everyone got what they wanted
- Was voting a good way? How would better ways be found? And evaluated?

Dissent!
More science? Scientists would say that
Sociology of science
China -> Tanzania

Recommendations
Gap map of theories
Theory hubs
Theory-focused synthesis
Putting theory in ToC
Theory-testing in evaluations
Think about markers