Evaluation Design 2: Methodological Choice

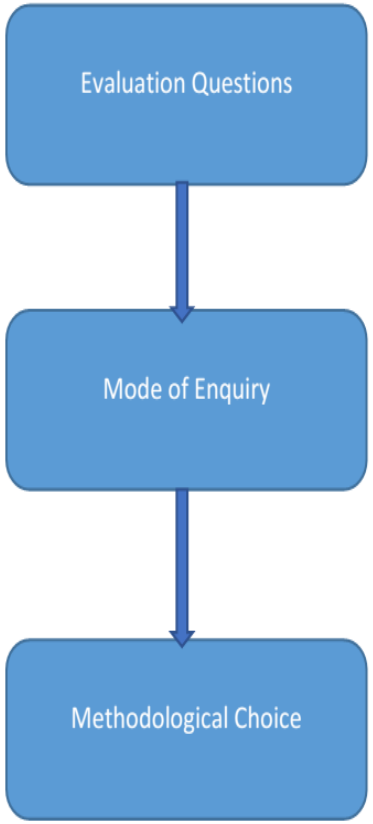
Elliot Stern

Evaluation Helpdesk Training

Brussels

April 17th-18th 2018

Linking EQs and Methodological Choice



In the previous presentation, Evaluation Questions were described as:

'a crucial link between evaluation purposes and how we make methodological choices'

Evaluation questions if they are well formulated point us to different 'Modes of Enquiry' which themselves suggest 'Methodological Choices'

Modes of enquiry

You will recall that some evaluation questions are ‘descriptive': *'Have programmes achieved their objectives?’*

Some EQs are ‘explanatory’ :*‘Can we demonstrate that programmes caused the results?’*

Some EQs are future-oriented: *‘Are these capacities likely to be self sustaining and economically viable?’*

Descriptive, explanatory and future oriented modes of enquiry draw on different families of methods – beware of evaluators that irrespective of EQs always offer the same ‘solution’ – the method they believe in!

An emphasis on results & impacts

- The current programming period sets out to strengthen the result-focus of EU programming (Article 56(3) CPR
- This is associated with encouragement for ‘impact evaluations’ that set out to examine *whether*, the *extent to which* and *how* a programme caused the intended effects.

Results and impacts

- This presentation therefore gives special attention to impact-related issues around evaluation design – even though the distinction between ‘impact’ and other forms of evaluation are not always that clear-cut!
- The EVALSED Guide gives a good summary of specific methods and debates and difficulties in the field of ‘impact evaluation’ – this presentation does not attempt to replicate that summary

Defining Impact Evaluations

Impact Evaluation (IE) in policy settings sets out to do three things:

- First to demonstrate that a programme '*caused' an 'effect'* – the intended results
- Second IEs are often expected to explain *how* a programme works
- Third to consider the *contribution* a programme makes

Approaches to Causal Inference: Counterfactual and Theory Based

- There are different approaches to explanation and ‘causal inference’
- The latest EVALSED Sourcebooks suggest that traditional counterfactual analysis can be good at demonstrating ‘cause’ and the extent of effects; whilst understanding ‘how’ and ‘why’ is better achieved with ‘theory-based’ evaluations

This is a reasonable practitioner-friendly position but the messages are clear……

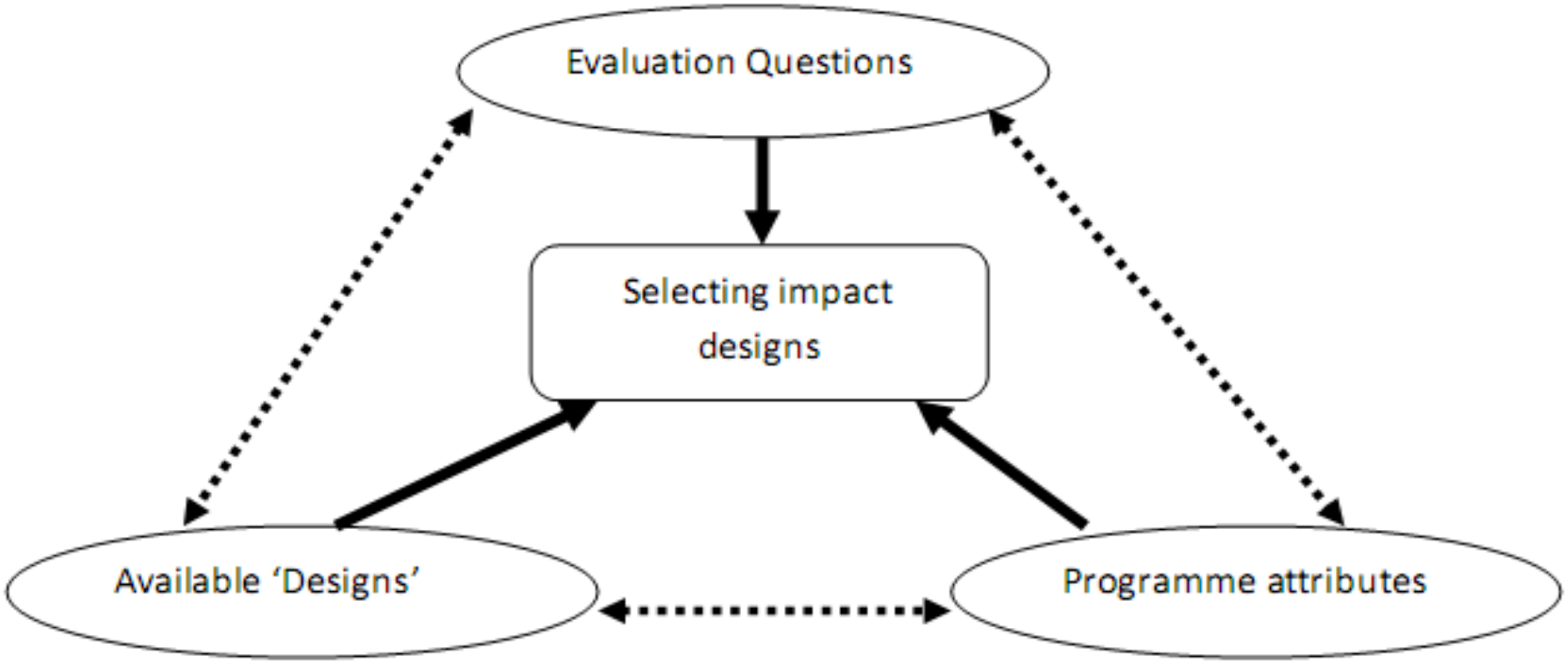
Rigorous Impact Evaluation is Difficult

- Top-end Impact Evaluations moves evaluation to the frontiers of sophisticated, expensive and still-developing research methods in the social and economic sciences
- Methods such as propensity score matching, discontinuity designs and instrumental variables (on the counterfactual side); and QCA, Contribution Analysis, Process Tracing and Realist Evaluation in relation to Theory Based Evaluations are not simple
- There are few specialists in most EU Member States able to apply them
- This is why we need to be cautious and selective about when we would expect to do this work

There are more modest ways of evaluating results

- Sometimes we can rely on ‘descriptive inference’ – an accumulation of different kinds of evidence about change occurring in some programmes but not in others without the same interventions
- If for accountability purposes we want to be able to say that a) there has been an improvement in results and b) that it seems to be aused by a programme then knowing *how* the programme worked (explanatory analysis) may not be needed
- Sometimes we can rely on prior evaluation, research and theory – we do not need to ‘prove’ again that smoking is bad for your health or that nutritional food is good for children! Here we might simply focus on implementation and uptake

The Design Triangle



Available designs

The ‘design triangle’ reminds us to balance our questions with available designs

- By available designs I mean the main ways in which it is possible to say anything about cause and effect using accepted and trustworthy approaches
- There are only a limited number of such approaches but there is more than one or two suitable in different circumstances – i.e. when we ask different EQs and when we deal with different kinds of programmes

Four Types of Causal and Explanatory Designs

- **Regularity frameworks** that depend on the frequency of association between cause and effect - the inference basis for statistical approaches to IE
- **Counterfactual frameworks** that depend on the difference between two otherwise identical cases – the inference basis for experimental and quasi experimental approaches to IE
- **Multiple causation** that depends on combinations of causes that lead to an effect - the inference basis for 'configurational' approaches to IE
- **Generative causation** that depends on identifying the 'mechanisms' that explain effects – the inference basis for 'theory based' and 'realist' approaches to IE

Strengths and weaknesses of different designs

'*Regularity*' requires high numbers of diverse cases to capture sufficient diversity (or difference) and numbers of variables for measurement

Counterfactuals are good at answering the question: 'Has this particular intervention made a difference here?' But weak on external validity questions: 'Will it work elsewhere?'

Generative causation is strong on explanation but weak on estimating quantities or extent of impact.

Experiments and *regularity/statistical association* approaches work best when causal factors are independent, but not if causal factors interact

Neither experiments nor statistical models are good at taking account of cultural, institutional, historical and economic settings but are good at measurement in the right circumstances…..

Combining Designs and Methods

- Most evaluations will need to adopt multiple designs to answer different evaluation questions

Rarely will any single design be sufficient

- We also combine different families of methods to build on their strengths and compensate for their weaknesses when answering the same question!

Methods as Representing Overall Designs

- I have avoided talking about specific methods or techniques so far and have concentrated on ‘designs’
- But methods can represent overall designs and families of methods

[HANDOUT 3/1]

Back to Programme characteristics

The Design Triangle reminds us that programme characteristics also shape evaluation design

This derives both from the programme's 'theory of change' and from other characteristics such as innovativeness, the complexity of the programme and how 'embedded' or isolated a programme is

Programme Characteristics and IE Design

Here are a few design implications to think about:

- An established intervention that is known to work reliably elsewhere, justifies an evaluation focus on preconditions and implementation but not on ‘whether it works’ – except through monitoring
- A programme with a number of intervention ‘strands’ likely to be implemented differently across sites should consider a ‘configurational’ approach such as QCA
- An programme that is moderately complex (several intervention strands) but where sites are embedded in a context that is likely to be influential, should consider ‘Realist’ evaluation approaches

Impact Evaluations are not just about Impact!

A results-orientation and IE should not be separated from other evaluation traditions. We continue to be interested in….

- How to *deliver* to achieve impact? – implementation or delivery evaluation
- How to *engage* with excluded groups? – participatory or local evaluations
- How to *build capacity* for monitoring? – capacity development
- How to *adapt to context* - demographic, institutional or cultural? – Needs analysis and comparative studies a programme's context

This is why evaluators who offer IE and seem to know little about other types of evaluation can be a risk

Returning to Evaluation Questions – Not all IE questions are the same

1. To what extent can a measurable impact be attributed to this intervention?

Experiments and statistical models

2. Did the intervention make a difference or contribute?

Process tracing and Contribution Analysis

3. How has the intervention made a difference?

Theory-based and structured case-based methods

4. Will the intervention work elsewhere/elsewhen?

Methods that take context seriously – Contribution analysis & Realist approaches

Limitations of the Single Evaluation

- Most of the designs and methods discussed so far have related to single cases
- Most Programmes include multiple cases – firms, communities, infrastructures and service providers; and every Member State have multiple programmes each containing multiple cases
- We know that learning and reliable conclusions through evaluation, also has to be built on multiple cases – what we call synthesis methods
- This raises questions about the unit of analysis for evaluations cross-programme? cross-OP? and how to plan integrated evaluation programmes

Ensuring the validity of what we evaluate and of our conclusions

- Evaluation and IE in particular is becoming increasingly technical
- However evaluation ultimately is about valuing – a matter of people applying their judgement criteria
- The validity of constructs (the things we evaluate) and of judgements will be challenged if they do not make sense to stakeholders and beneficiaries – this is the classic criticism of indicator-based monitoring
- A good IE – and most other evaluations – has to ensure that the core concepts, and what we measure and describe makes sense in the real world
- The same applies to when conclusions and recommendations are made

Engaging stakeholders, beneficiaries and others implicated in evaluations

Validity is a major 'quality' criteria in any evaluation

- Systematically involving users, stakeholders, beneficiaries, panels of citizens, policy makers is a good way to improve validity
- This needs to take place at every stage in the evaluation process – from operationalising criteria – e.g. deciding what outcome descriptor or measure represents 'success'; to interpreting data when drawing conclusions and making recommendations

These things are too important to leave to evaluators!

Conclusions

- Evaluations need to be designed to take account of Evaluation Questions; the characteristics of programmes; and the capabilities of available methods
- Evaluations often use different designs to answer different EQs; & use different methods because all have their strengths and weaknesses
- IEs variously need to establish causality; explain why and how; and to assess how programmes interact with other causative factors- however IEs and results are not just about causl & explanatory designs
- Deciding on the ‘unit of analysis’ for an evaluation – sites, programmes, multiple programmes - is also part of the design process
- Validity will be increased if users, stakeholders, beneficiaries and citizens are involved in evaluative design and evaluative judgements

Exercise – Evaluation Design

- In your ‘pack for this Block there is a checklist that can be used to assess proposals for an Impact Evaluation. It covers many of the ‘design’ issues raised this morning.
- Please consider this checklist in relation to a programme you know and can share with the colleagues in your group
- Spend 20 minutes on this and then you will have an opportunity to feedback your thoughts and reactions