

# Theory-Based and Impact Evaluation

# WELCOME!

## *Purpose of these 2 days*

- To broaden your understanding of 'Theory Based Impact Evaluation' (TBIE)
  - To introduce you to core ideas in Impact Evaluation (IE) more generally
  - To increase your ability to commission and manage TBIEs
- Suggest frameworks and checklists that can help you assess and plan IEs

*It is not to make you into an TBIE expert but rather to help you have sensible conversations with experts who you may work with – although we cannot avoid technical content entirely....*

# Commission and Manage Impact Evaluations

This course is intended to help you when...

- Drawing up IE terms of reference
- Assessing IE proposals that cross your desks
  - Managing and steering ongoing IEs
- Assessing the strength of conclusions and recommendations reached by those conducting IEs
- Developing new programmes and policies that are ‘evidence-based’, i.e., learn lessons from completed IEs

*To do these things well you do need background understandings*

# Form and Format

Mixture of activities that include

- Presentations
- Group discussions
- Questions and answers
  - Exercises

# Part One: What is Impact Evaluation & Theory Based Impact Evaluation ?

Why it matters, where it comes from and how to define it

# An emphasis on results & impacts

- Recent programming periods sets out to strengthen the result-focus of EU programming
  - In the current Programming Period Managing Authorities are expected to prepare Evaluation Plans and to ‘provide a framework to plan *impact* evaluations’ (Article 56(3) CPR)
  - ‘..... the managing authority shall carry out an evaluation for each programme to assess its *impact* by 30 June 2029’

# An emphasis on results & impacts

This emphasis on results & impact highlights the *purposes* of evaluation

- In the early stages of Structural Fund evaluation we were mainly interested in measuring success and failure – were objectives met, how many kilometres of roads were built, how many technicians were trained?

Now purposes have expanded....

# Evaluation Purposes

*At ex ante stage*

- Problem analysis and Planning

*At Ongoing and ex post stages:*

- *Accounting for Results*
- *Better Management and Delivery*
  - *Learning and Improvement*
- Capacity development and strengthening

*This emphasis on accounting for results and learning so as to improve is what drives the demands for impact evaluation*



# Results and impacts

- The EVALSED Guide summarises a number of specific methods, methodological debates and difficulties in the field of IE
- This presentation does not attempt to replicate that summary although we will be looking inside the *logic* of various methods

## *Evaluating Results and Impacts*

What should we be looking at?

- IE is about cause and effect

*Can we demonstrate that the programme caused the intended outcomes?*

You can never observe causality: as with all evaluation, judgements have to be made on the basis of evidence and logical arguments

This is the reason we speak of 'causal inference' or 'making a causal claim' – of which more later

## *Evaluating Results and Impacts*

But this is not straightforward - we need to know:

- If the economy improved would unemployment have gone down anyhow? Did the grant go to already successful SMEs? Was the increases in innovation the result of previous investments in R&D; was city-centre renewal caused by a new retail park or something else

IE sets a much higher standard of 'proof' that results were really *caused* by the programme; and arguing convincingly even on the basis of good evidence can be difficult

## *Evaluating Results and Impacts*

- For improvement purposes, we need to *explain* – how and why did an effect happen?
- In order to improve we need to understand what goes right and what goes wrong but *how* and *why* things go right and wrong

Was the failure of an innovation programme because of poor regional research facilities? Poor implementation? Uncooperative trade associations? Bad coordination and timing? Unrealistic goals? An economic crisis...

## *Evaluating Results and Impacts*

We also know that in a complex world causes are not singular

- An effect is often the result of more than one cause
- Policy interventions interact with other non-policy factors
  - Policies interact with other policies

So we also want to know about the *contribution* of our policies

# Defining Impact Evaluations

To summarise: 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 expected to *explain how* a programme works
  - Third IEs consider the *contribution* a programme makes

Question and Answer Point

## *Evaluating Results and Impacts*

If IE is understood to be about causal attribution, explanation *and* contribution it requires methods that can do this job!

- The Commission's guidance distinguishes between 'experimental' or counterfactual IEs, *and* what are called 'Theory-Based Impact Evaluation' both recognised in Commission Guidance
- The latest EVALSED Sourcebooks suggest that 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

A bit simplistic but .....

## *What is Theory Based Evaluation?*

In the past evaluation avoided theory!

- Achievements of success/failure was judged against standards and criteria – achieving objectives the main goal of evaluation
- This changed with a growing interest in using evaluation to guide improvement and change – when evaluation and policy collided
- We needed theory and to say something about what *causes* an observed or desired effect



## *What is Theory Based Evaluation?*

### Different meanings of 'theory'

It can mean:

- Accumulated knowledge and experience – practice theory
- Hunches we have about how the world works – verging on beliefs
- Verified results of systematic investigation – e.g. from evaluation
- Rules of logic, methodology & explanation- 'theories of knowledge'
- Domain theory: how innovation, social cohesion & competitiveness works

*All of these forms of theory inform evaluation but theories of knowledge are most central to TBIE and associated methodologies*

## *What is Theory Based Evaluation?*

- There are different ‘brands’ of theory-based and theory-informed evaluations: Theories of Change; Realist Evaluation; Contribution analysis are the better known ‘brands’ – of which more later
- Overall TBE is method neutral: you can potentially use a wide range of quantitative and qualitative methods within TBE – statistical modelling; ethnography; interviews; observation studies
- What characterises TBIE is the way causal claims are justified – the overall logic and the overall design of the evaluation

## *What is Theory Based Evaluation?*

- TBIE claims to open up the ‘black-box’ – get under the surface of cause and effect – most scientists would agree that you cannot observe causality – you have to make assumptions, have a theory in order to make a causal claim
- The difference is how theory helps to attribute causality and explain ‘how’ and ‘why’

*A key job for evaluation is to identify, construct or reconstruct the ‘theory’ of a policy, programme or project*

## *What is Theory Based Evaluation?*

### Where does 'theory' come from?

- From the assumptions of policy makers – e.g. 'programme theory' or 'intervention logic' or 'logic models'- 'to achieve competitiveness I must support exporters'
  - From the assumptions of multiple stakeholders – e.g. residents, entrepreneurs and municipal leaders – often the job of the evaluator is to differentiate, understand and even build consensus
- From knowledge produced through academic research e.g. how new research is taken-up by enterprises and leads to economic growth
  - From methodologists and philosophers of science: e.g. realists, positivists or social constructionists – how can we interpret evidence

# All Rigorous Impact Evaluation is Difficult

- We also need to be cautious and strategic about when considering any kind of Impact Evaluation
  - Top-end Impact Evaluations moves evaluation to the frontiers of sophisticated and still-developing research methods in the social and economic sciences
    - Methods are not simple and they are expensive
  - Few specialists in most EU Member States are able to apply them
- We must therefore be cautious and selective about when we would expect to do this work

Question and Answer Point

Group Discussion

# Part Two: Evaluation Design

Why design matters: How Evaluation Questions, Methodologies and Programme Architecture shape Impact Evaluation,

# Designing Evaluations

Evaluations need to be *designed*

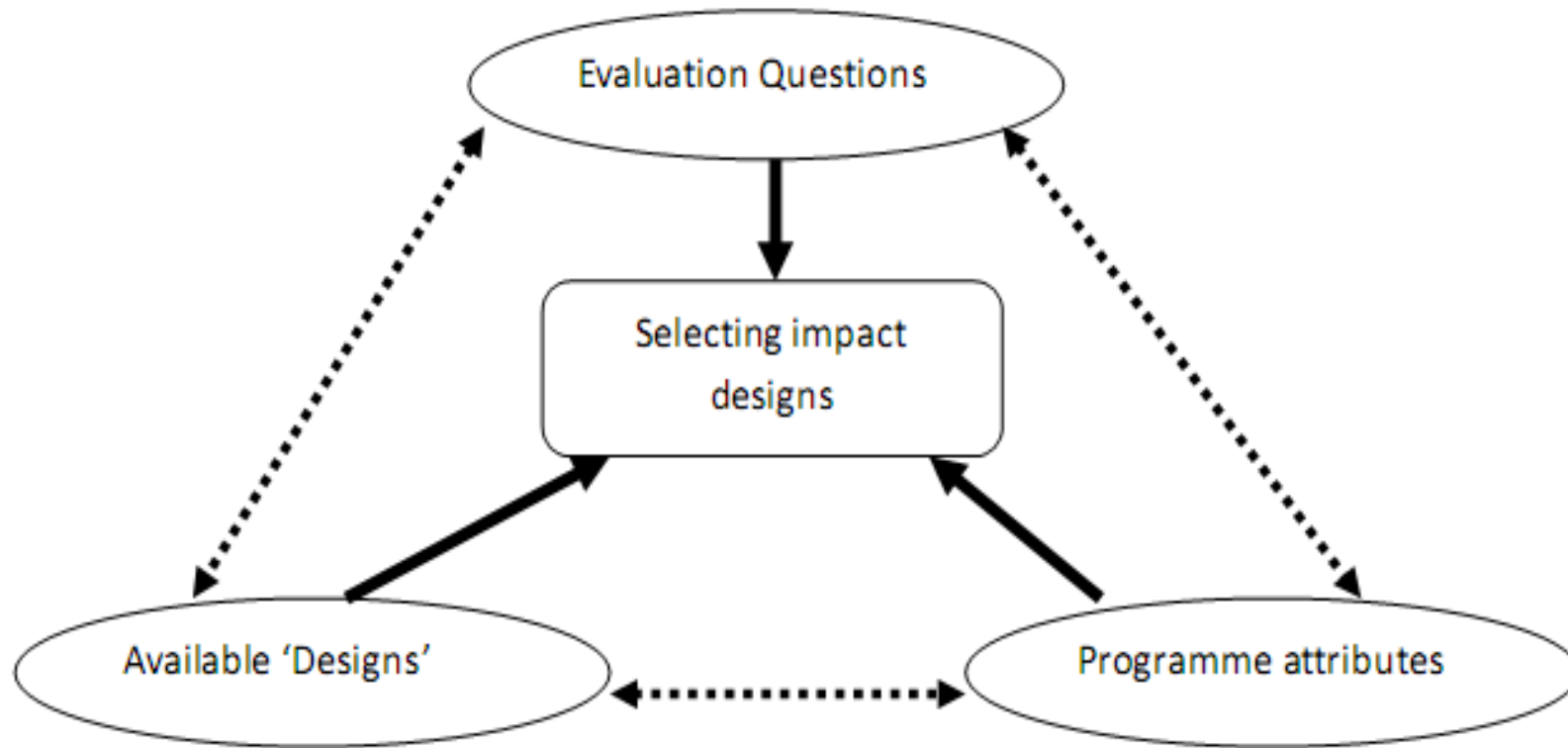
- This starts with an ITT or specification leading to a proposal for an appropriate methodology
- Weak evaluations nearly always start with a poor ITT and a badly thought out methodology
- As part of work of the helpdesk, we have reviewed the quality of many Structural Fund evaluations
- Poor quality evaluations were often traced back to weak designs

# Designing Evaluations

- A 'design' is more than a method
- Interview surveys, case-studies, statistical analyses of labour market data, observational studies, all can be part of different designs
- A design involves a deep understanding of what we want to know, the programme context in which we are working and of the capabilities of different families of methods
- On the basis of this understanding we can consciously choose which of the many ways we could evaluate any programme



# The Design Triangle



# Designing Evaluations

The 'Design Triangle' suggests we need to match:

- Evaluation Questions (what we want to know )

*with the*

- Available designs and their capabilities (what these designs can do given the Evaluation Questions being asked and the characteristics of Programmes)

*with the*

- Characteristics of Programmes (the 'object' to be evaluated)

# Importance of Evaluation Questions

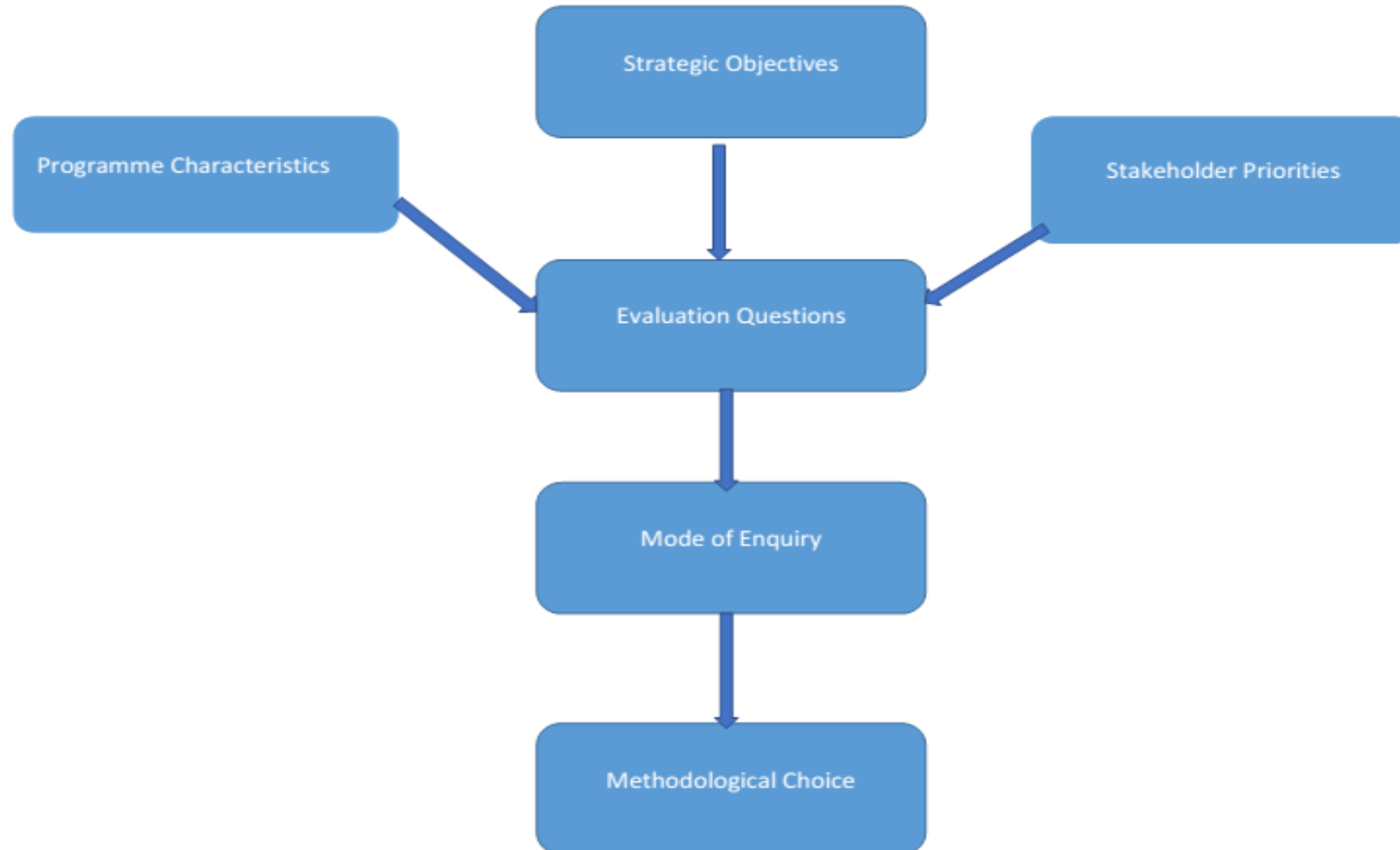
- Historically evaluations were structured around generic criteria such as Effectiveness; Relevance; Efficiency; Coherence; and Sustainability
- These criteria-led evaluations often produced evaluations that lacked specifics and were not useful for policy purposes. Nowadays we tend to operationalise criteria into more specific *Evaluation Questions*
- Evaluation Questions (EQs) are also a crucial link between evaluation purpose and how we make methodological choices
  - EQs have methodological consequences

*Evaluations need to answer clear and answerable questions!*

# Importance of Evaluation Questions

- EQs are not the same as an ‘interview question’ – they are ‘high-level’ question able to be analysed and studied
- EQs focus on the relationship between a programme and real world ‘consequences’ ‘effects’ and ‘results’
- EQs that only focus on the internal operations of a programme and do not try to illuminate the relationships between programmes and real-world effects are not EQs!
  - Most evaluations will have a number of prioritised EQs – each addressing a specific point of interest in a programme

# Centrality of Evaluation Questions



# Evaluation Questions and Strategic Objectives

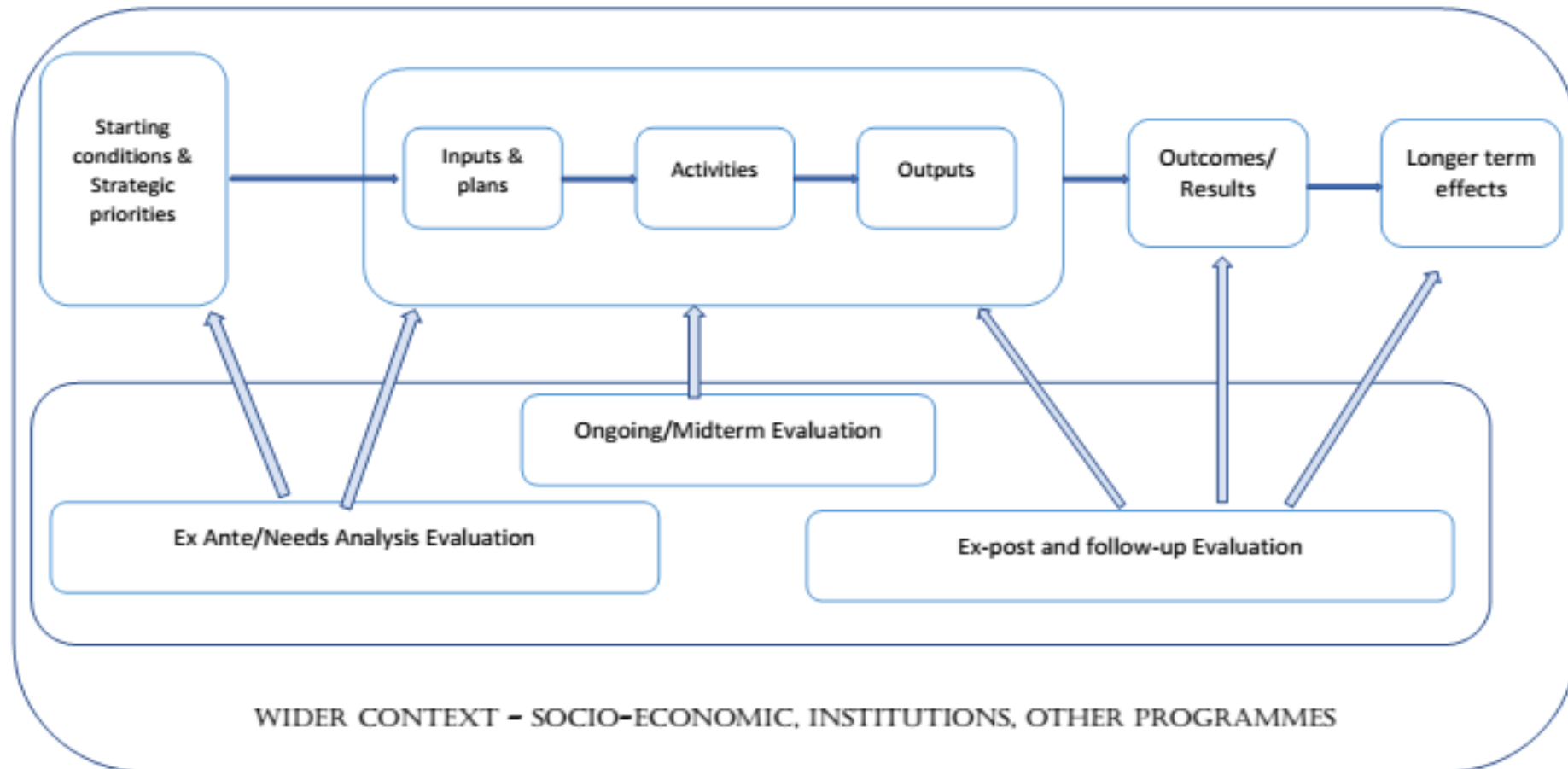
- Strategic Objectives refers to the policy goals that justify a programme – in our case through the various programme priorities – strengthening research; access to ICTs; enhancing competitiveness; promoting social inclusion etc. - that are intended to contribute to ‘smart, sustainable and inclusive growth’
- Evaluation Questions need to relate to these goals and priorities although monitoring/indicator exercises that are not necessarily evaluative will also address results

# Programme Characteristics

Evaluation Questions have to take account of programme characteristics

- There are many ways to characterise a programme – in terms of sectors, objectives, their degree of innovativeness etc.
- For the purpose of specifying EQs, characteristics should first be understood in terms of an ‘intervention logic’, ‘theory of change’ or ‘programme theory’ set into a wider context
- Simple Theories of Change usually begin from an intervention-logic - the sequencing of a programmes cycle set into a wider socio-economic and regional context

# Programme Characteristics





# Programme Characteristics

When we discuss methodological choices there are other kinds of programme characteristics that also have to be considered. These include for example:

- Innovativeness of programme goals and delivery
  - Simplicity or complexity of the intervention
- Types of outputs envisaged – whether they are material, behavioural, new services, new institutional arrangements
  - Timescales and trajectories of change
- How bounded or embeddedness programmes are in relation to other programmes, activities and systems

# Attributes also have design implications

- Overlap with other interventions with similar aims different activities and projects in one place – multiple causal factors
- Customised non standard interventions adjusted to different contexts – have to assess what works in what circumstances – Realist approaches do this
- Programmes working ‘indirectly’ through ‘agents’ - so we need more than one Logic Model/Theory of Change/ Causal pathway
- If likely impacts of programmes are long term this influences a key design decision: when to conduct an IE and follow up
- Intended impacts are difficult to measure and in parts intangible – requires qualitative as well as quantitative approaches

# Stakeholder priorities

- Evaluations like programmes have stakeholders
- Stakeholders will want answers to *their* questions
- Stakeholders are the users of evaluation – they are also often the gatekeepers to evaluation data and provide necessary cooperation for evaluators
- The credibility and legitimacy of an evaluation will often depend on how far the questions of stakeholders are prioritised

Stakeholders underpin 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, values and 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 has to ensure that the core concepts, and what we measure and describe makes sense in the real world at the design stage and when making conclusions and recommendations

Question and Answer Point

# Four types of Impact Evaluation Questions

1. To what extent can a measurable impact be attributed to this intervention?
2. Did the intervention make a difference or contribute or add-value?
  3. *How* has the intervention made a difference?
  4. Will the intervention work elsewhere/elsewhen?

# Available designs: Approaches to Causal Inference

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 answer questions about cause and effect using accepted and trustworthy approaches
- There are only a limited number of such broad approaches but there may more than one or two suitable depending on different circumstances

# Four Types of Causal Approaches

- **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

## *All approaches have strengths and weaknesses*

<b>Requirements, Strengths and Weaknesses of Four Approaches to Causal Inference</b>				
	<b>Regularity</b>	<b>Experiments/ Counterfactuals</b>	<b>Multiple causation</b>	<b>Generative/mechanisms</b>
<b><i>Requirements</i></b>	Many (or highly diverse) cases Independent causes	Two identical cases for comparison Ability to 'control' the intervention	Sufficient number of cases Availability of cases with comparable characteristics	One case with good access to multiple data sources Theory to identify 'supporting factors'
<b><i>Potential Strengths</i></b>	Uncovering "laws"	Avoiding bias	Discovery of typologies Dealing with limited complexity	In-depth understanding Focus on the role of contexts 'Fine-grained' explanation
<b><i>Potential weaknesses</i></b>	Difficulties explaining 'how' and 'why' Construct validity	Generalisation Role of contexts	Difficulties interpreting <u>highly</u> complex combinations	Estimating extent Risks of bias/ loss of evaluators' independence



# 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 dealing with contextualisation – taking account of cultural, institutional, historical and economic settings but are usually good at measurement .....

# Design Approaches & Causal Inference

Theory-based: Causal chains	<i>Causal process designs</i> : Theory of Change, Process tracing, Contribution Analysis, impact pathways, <i>Causal mechanism designs</i> : Realist evaluation, Congruence analysis	Identification/confirmation of causal processes or 'chains', Supporting factors and mechanisms at work in context
Theory based: 'Case-based' approaches	<i>Interpretative</i> : Naturalistic, Grounded theory, Ethnography <i>Structured Configurations</i> : QCA, Within-Case- Analysis, Simulations and network analysis	Comparison across and within cases of combinations of causal factors Analytic generalisation based on theory

Question and Answer Point

Discuss Handout 1

# Combining Designs

- Most evaluations of effects or impacts will need to adopt multiple designs and methods to answer their evaluation questions

*Rarely will any single design be sufficient*

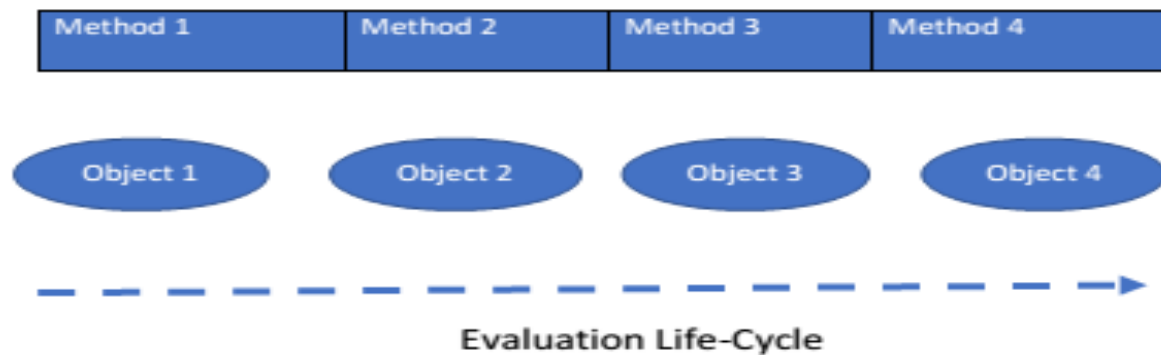
- Most good IEs will use several designs to complement each other – to build on their strengths and compensate for their weaknesses!

# Combining Designs

Triangulation



Multi-method Designs



# When to use RCTs on their own?

- You are interested in a single cause and single effect
- You are dealing with a standardised not customised intervention
  - You have sufficiently large numbers for statistical purposes
- You can precisely match and control the population and a comparison group
  - You want to understand what *has* happened in a particular place or time
- You are *not* primarily interested in explanation or generalisation

# EQs have different methodological implications

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 ToC/Contribution Analysis*

3. How has the intervention made a difference?

*Theory-based and comparative 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

# Design Approaches & Causal Inference

Question and Answer Point

Discuss Handout 2