

cedilprogramme.org



# Evaluating complex interventions

Edoardo Masset

Mexico City

14 October 2019

# CEDIL Programme of work 1: unpacking complex interventions

- CEDIL methods studies
- Commissioned academic papers
- Impact evaluations of complex interventions

# What are complex interventions?

## Complex interventions

Interventions with: multiple interacting components and embedded in complex systems

- Path dependence
- Phase transition/multiple equilibria
- Feed back loops
- Tipping points
- Emergent properties

## Complicated interventions

Interventions with:

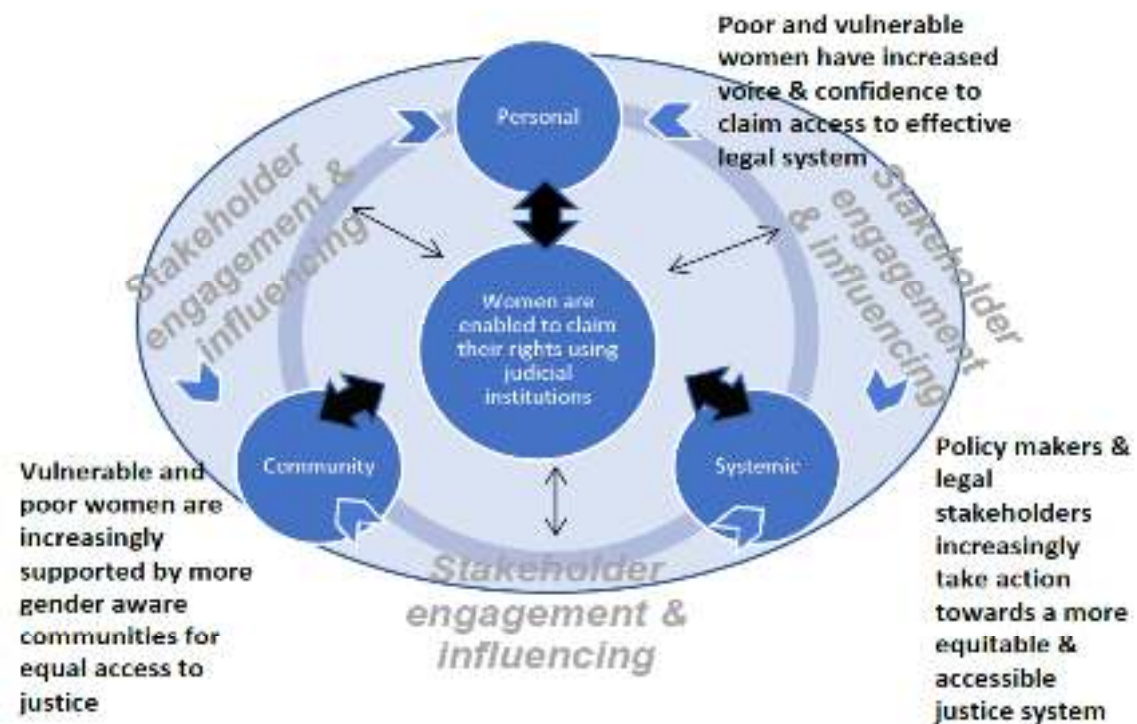
- Multiple components
- Multiple behavioural assumptions
- Multiple targets and implementers
- Multiple outcomes

**Multiple interacting components** with emerging outcomes. Example big push poverty eradication programmes

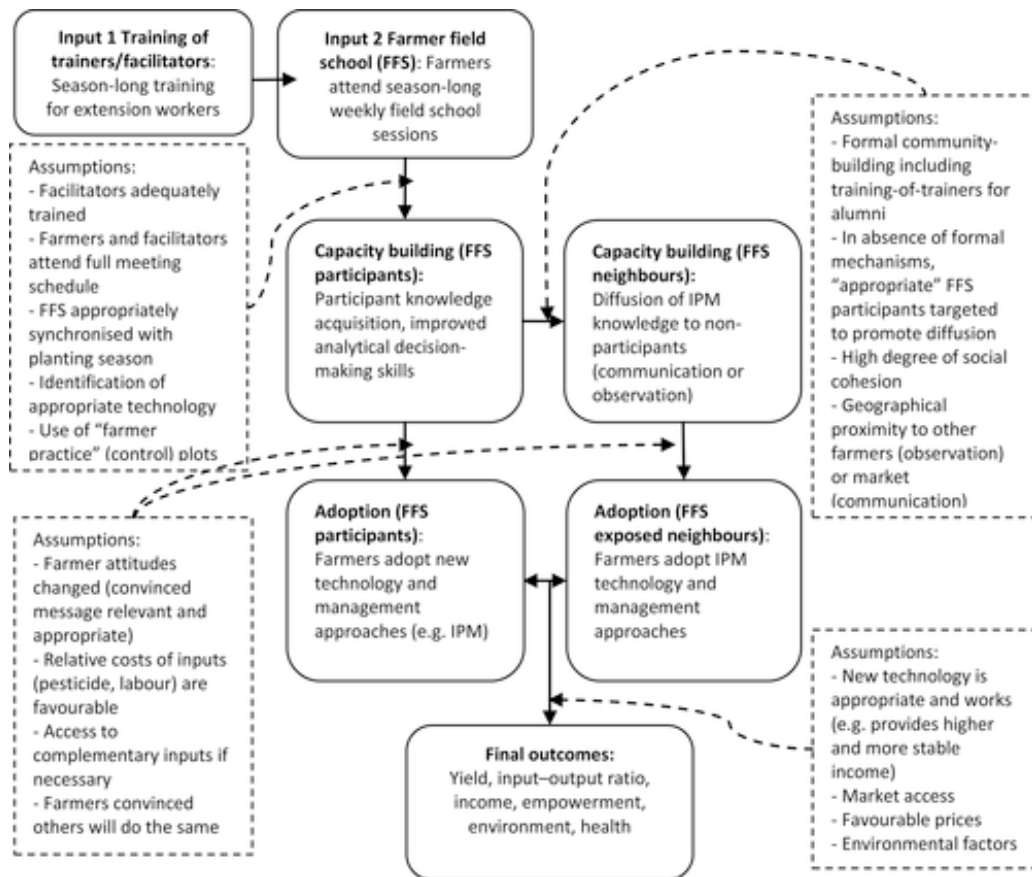


Many target populations and stakeholders at different levels

Figure 2.1: Theory of Change of the WAJ2 project



Women's Empowerment in Lebanon: Impact evaluation of the project Women's Access to Justice in Lebanon. Effectiveness Review Series 2017-18



Many behavioural assumptions and long causal chains

Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review, Waddington et al. (2014), Campbell Collaboration



## **Portfolios** interventions:

- similar interventions in many countries
- many interventions in the same country

Example Feed the future: nutrition and agriculture interventions in 12 countries.

# Approaches to evaluating complicated interventions

- Complex evaluations are rarely evaluated
- **Multi-arm or multi-site evaluations** are difficult or too expensive
- **Black box evaluations:** evaluate the impact of the 'package'
- **One-component-at a time evaluation:** evaluates impact of just one component or causal link



# Problems of standard approaches

- Standard approaches isolate the project effect after controlling for everything else
- In complex interventions there are many 'projects' and they interact with each other
- Impacts vary with other interventions and with the context
- Interest is to estimate how impact varies

# How to explore interactions and heterogeneity

- Design approaches: experimental tradition, factorial designs and adaptive trials
- Analysis approaches: structural modelling and DAGs
- Qualitative approaches: qualitative comparative analysis and process tracing

# Factorial designs

- Subjects or clusters are randomly allocated to treatments and combinations of treatments
- There is no 'business as usual' or 'no intervention' control group
- The control group is an average of all treatment combinations
- Goal is not testing hypotheses, but to screen factors for a multi-component intervention



# Example: smoking cessation

Piper et al. (2015), Identifying effective intervention components for smoking cessation, *Addiction*, 111

- 650 smokers willing to quit are recruited
- 6 treatment combinations are randomly allocated:
  - Nicotine patch
  - Nicotine gum
  - Counselling
  - Intensive counselling
  - Phone counselling
  - Intensive replacement therapy
- Results:
  - Intensive counselling effective
  - Positive interaction (synergy) between nicotine patch (or gum) and intensive counselling
  - Negative interaction of intensive counselling and phone counselling



runs	1	2	3	4	5	6
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	1	1	0	0	0	0
5	0	0	1	0	0	0
6	1	0	1	0	0	0
7	0	1	1	0	0	0
8	1	1	1	0	0	0
9	0	0	0	1	0	0
...	...	...	...	...	...	...
64	1	1	1	1	1	1

# Factorial designs

## ADVANTAGES

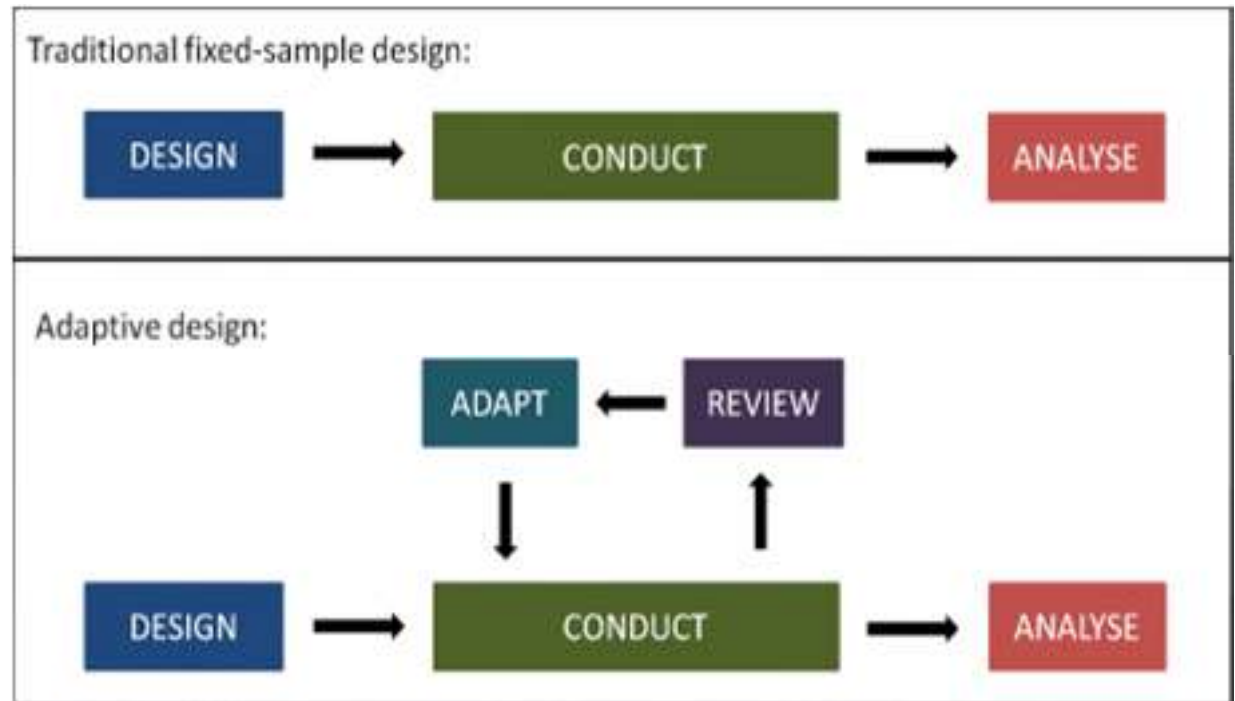
- Uses small samples
- Estimate interaction effects
- Finds 'critical mixes'

## DISADVANTAGES

- Inference is not strong and needs validation
- No direct comparison with a control group
- Difficult to implement in the field

# Adaptive trials

- Changes are pre-planned
- Possible changes:
  - Sample size
  - Allocation to treatment
  - Stop treatment or entire trial
  - Changes in hypotheses tested



Adaptive designs in clinical trials: why use them, and how to run and report them, Pallman P. et al (2018), BMC Medicine

# Example: HIV treatment

Pushpakom et al. (2015) BMJ Open

- A drug was administered in combination with HIV anti-retroviral therapy
- 4 different doses were experimented on 350 patients:
  - No drug
  - 20mg
  - 40mg
  - 80mg
- After 24 weeks an interim analysis was conducted
- Two low-dose arms were stopped for futility, the high does was continued



# Adaptive trials

## ADVANTAGES

- Avoid futile testing
- Cost-effective approach to project design
- Allows testing of multiple treatments and interactions

## DISADVANTAGES

- Impacts must be observed in the short run
- Large risk of false positive and false negatives and resulting upward bias



# Structural modelling

- Models of behavioural relations between factors and characteristics
- Can be conducted alongside RCT to recover causal parameters and to validate predictions
- They estimate the impact of the interventions as other characteristics and interventions changes

# Example: education policies in Mexico

## PROGRESA programme

Attanasio O, Meghir C and Santiago A (2011) 'Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa.' *The Review of Economic Studies* 79(1): 37-66



PETER BATE—ID/B/110

- The structural model was able to replicate the experimental results
- There were general equilibrium effects: the project increased the wage rate thus dampening the impact of the programme but not by much
- Further simulations showed that targeting secondary school children would increase enrolment more

# Structural modelling

## ADVANTAGES

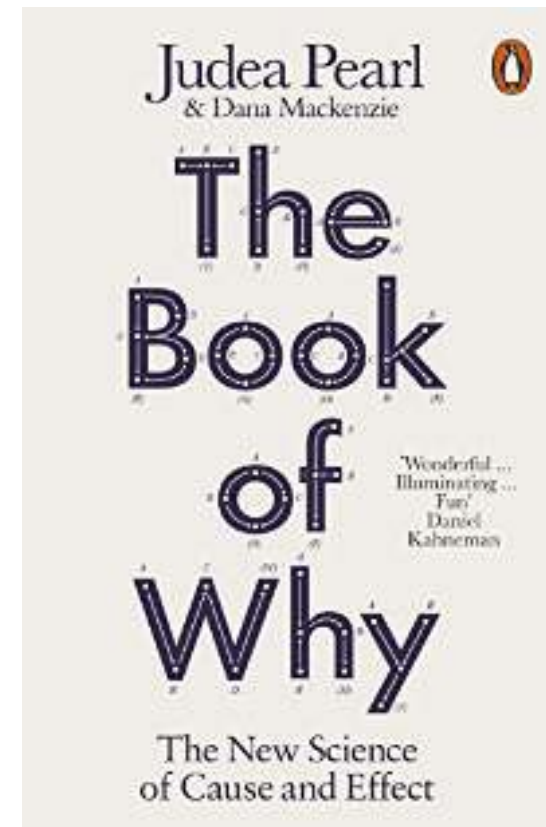
- Estimate impacts under different scenarios
- Estimate impact of hypothetical interventions

## DISADVANTAGES

- Complicated interventions require complicated models
- Dependent on assumptions
- With many factors parameters are difficult to identify

# DAGs: Directed acyclic graphs

- Proposes the **front-door criterion**
- An instrumental variable which is the mediator between the intervention and the outcome
- Strong causal inference without a control group

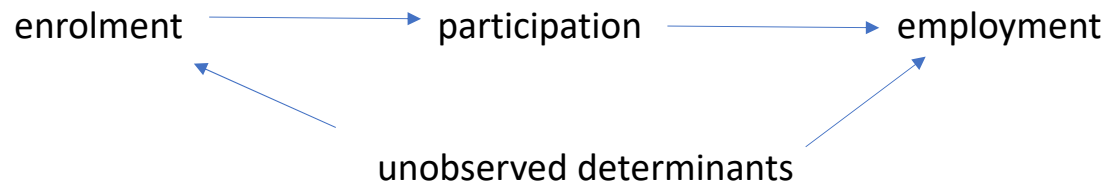


# Example: impact of the US job training programme

Glynn et al. (2017), Front-Door Versus Back-Door Adjustment With Unmeasured Confounding: Bias Formulas for Front-Door and Hybrid Adjustments With Application to a Job Training Program, *Journal of the American Statistical Association*, 113(523)

- The study replicates the results of the famous US national job training programme

## Front-door method:



- Strong assumption of no unobservables determining participation and employment
- There are no other applications

# Qualitative comparative analysis

- Multicausality: effects are produced by combination of causes
- Goal is identifying winning combinations of factors
- Does not compute effect sizes
- An algorithm that eliminates redundant combinations

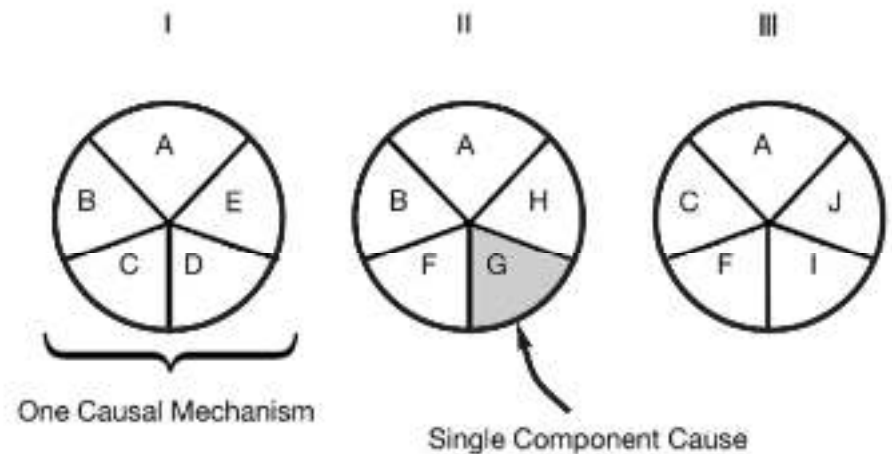


FIGURE 1—Three sufficient causes of disease.

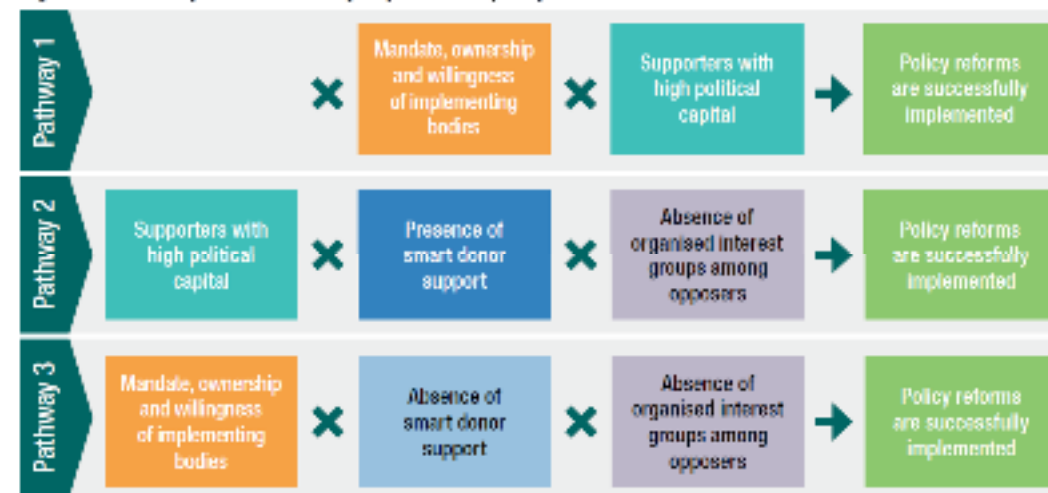
Rothman and Greenland (2005), *Causation and causal inference in epidemiology*, AJPM 95(1).

# Example: what drives policy change in Nepal?

Pasanen et al. (2019), What drives policy change in Nepal? A qualitative comparative analysis, ODI Report

- 18 case study of policy change in Nepal
- Four factors:
  - Willingness of implementing bodies
  - Supporters with high political capital
  - Smart donor support
  - Absence of organised opponents
- No single factor was necessary to success
- Some combinations of factors were successful

Figure 1 Pathways to successfully implemented policy reforms in the four-condition model 1



Note: x = in conjunction with.

# Qualitative comparative analysis

## ADVANTAGES

- Multicausality: impact of combinations of interventions
- Control group not needed
- Small sample size

## DISADVANTAGES

- Not simple: choices and deep case knowledge
- No statistical testing
- Not obvious interpretation of results



# Process tracing

- Within-case analysis: only one observation
- An outcome is observed
- A causal mechanism explaining the outcome is fully specified with a ToC/structural model
- Look for evidence in support of the mechanism and of alternative plausible mechanisms

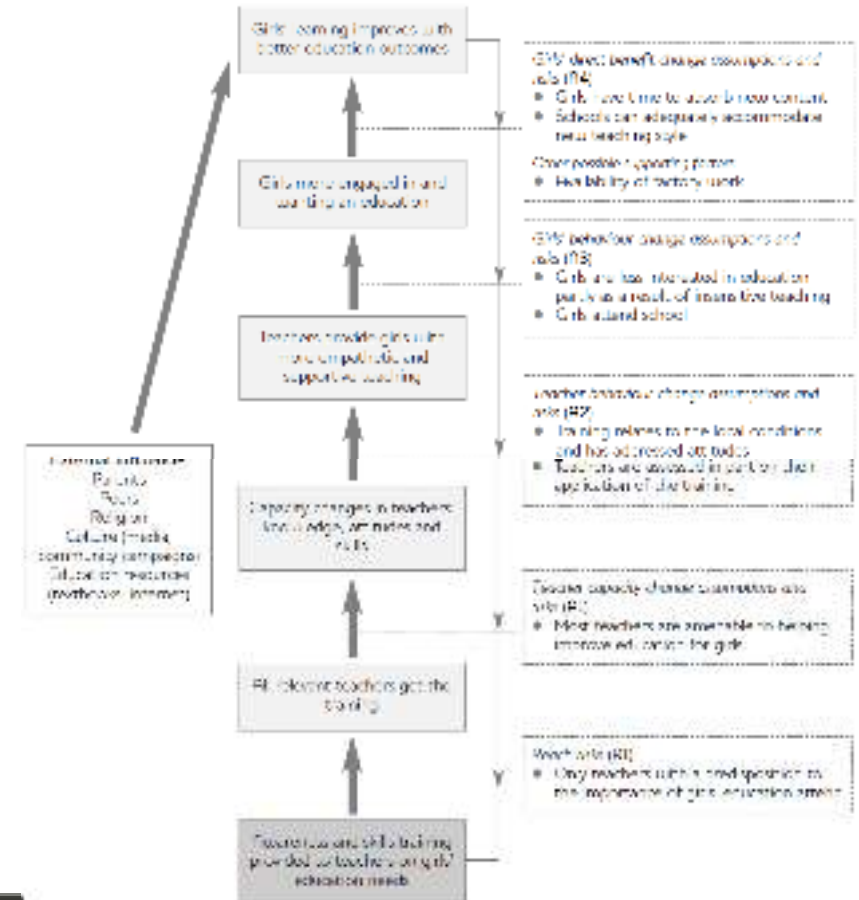


# Example: evaluation of a gender-sensitive education intervention

Befani et al. (2014), Process tracing and contribution analysis: a combined approach for impact evaluation, IDS Bulletin, 45(16)

- Promotion of gender-sensitive education is expected to promote girls' enrolment
- A positive impact is observed
- A full TOC is specified
- Two hypotheses are made:
  - a) teachers' training improved attendance
  - b) Closure of a factory increased enrolment
- Evidence is sought both in support and against each hypothesis
- Enrolment is found extremely unlikely under any other causal mechanism than teachers' training

Figure 2 ToC for enhancing education outcomes for girls



# Process tracing

## ADVANTAGES

- No control group
- No data collection
- Good understanding of causal mechanism

## DISADVANTAGES

- Fully retrospective
- Not externally valid
- No evaluation of interactions or multiple causes

# Alternative methods for complex evaluations?

- Are these methods appropriate to evaluate complex interventions?
- What is your experience in evaluating complex interventions?
- Are there other methods that we should consider?



## **CONTACT CEDIL**

cedilprogramme.org

E: [cedil@cedilprogramme.org](mailto:cedil@cedilprogramme.org)

T: +44(0) 20 7958 8136

🐦: [@CEDILprogramme](https://twitter.com/CEDILprogramme)

🌐: [cedilprogramme](https://www.linkedin.com/company/cedilprogramme)