Causal Pluralism and Mixed Methods in Impact Assessment

Paul Shaffer

Dept. of International Development Studies
Trent University, Canada

Workshop on Poverty, Inequality Dynamics and Economic Development: Tensions and Trade-offs in Mixed Methods Research

King's College, London

Sept. 6-7, 2018

Topics

1. Two Approaches to Impact Assessment
2. Two Approaches to Causation
3. Methodological Differences
4. Empirical Examples
5. A Case for Mixed Methods (Q2) in Impact Assessment (with caveats)

1. Two Approaches to Impact Assessment

Impact assessment – making a causal link between activities/outputs of programs, projects, policies and outcome or impact indicators

Experiments/Quasi-Experiments vs. Typical 'Qualitative' Approaches

Experiments:

Assignment to treatment and control group is random (RCTs)

Properties of randomness imply that treatment and control are identical in expectation (over large enough numbers)

Assumes that assignment is actually random, etc.

Need to deal with contamination, spill-overs, incomplete compliance, etc.

Quasi-Experiments:

Approximate experiments

Treatment and comparison groups created statistically based on observable characteristics or using other criteria (those just below and above a program eligibility threshold)

Examples include propensity score matching (using probit regressions), regression discontinuity designs, etc.

1. Two Approaches to Impact Assessment

Typical 'Qualitative' Approaches

Use a range of dialogical techniques (focus groups, semi-structured interviews, life histories) and others such as causal mapping, ranking, etc.

Core objective is to trace out linkages between programs/policies and outcomes or impacts, drawing on lived experience of individuals/communities and local understandings of drivers of change

At times, approach may ask about the magnitude of impact in broad, ordinal categories

Sometimes called 'process tracing'.

Caveats

Stylised Depiction of both approaches

Ignoring differences within 'qualitative' traditions (e.g. dialogical approaches vs. participant observation)

Hybridity/overlap possible (as shown in Q2 Examples)

2. Two Approaches to Causation

Many different concepts of causation and strategies of causal inference:
'One Word, Many Things' (Nancy Cartwright)

"nomological subsumption, statistical correlation, counterfactual dependence, agential manipulability, contiguous change, energy flow, physical processes and property transference" (Schaffer, 2008)

"Bayes-nets accounts and Granger causality, modularity accounts, manipulation accounts, invariance accounts and causal process theories" (Cartwright, 2007)

Difference-makers vs. Producers

Producers - "Causation is Physical Producing"
Difference-makers – "Causation is making a Difference"

2. Two Approaches to Causation

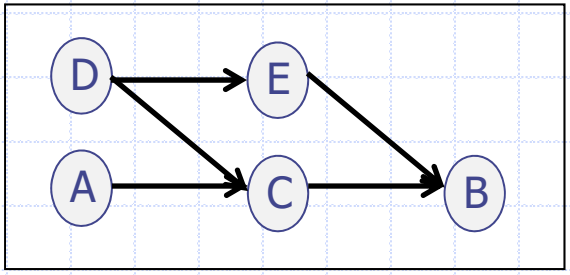
Producers - "Causation is physical producing"

"To assert that A's are cause of B's is to assert that there is a typical causal mechanism through which events or actions of type A lead to B."

- Causal mechanisms – no consensus definition
- In 'qualitative' impact assessment includes: i) causal variables; ii) causal tree; ii) some explanation of how variables do their causal work

How to identify causal mechanisms?

- Theoretically (theory-based evaluation)
- Empirically, using dialogical techniques: 'what did you do with the funds you borrowed', causal maps (next slide), etc.
- Ex. Autopsy (theory of signed causes)



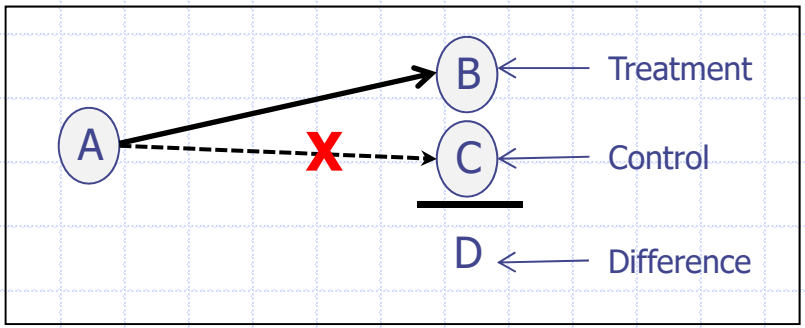
2. Two Approaches to Causation

Difference-makers – "Causation is making a Difference"

E.g.: Counterfactual Dependence – Event/Action A causes B if B would not have occurred in the absence of A. 'A makes a difference to B'

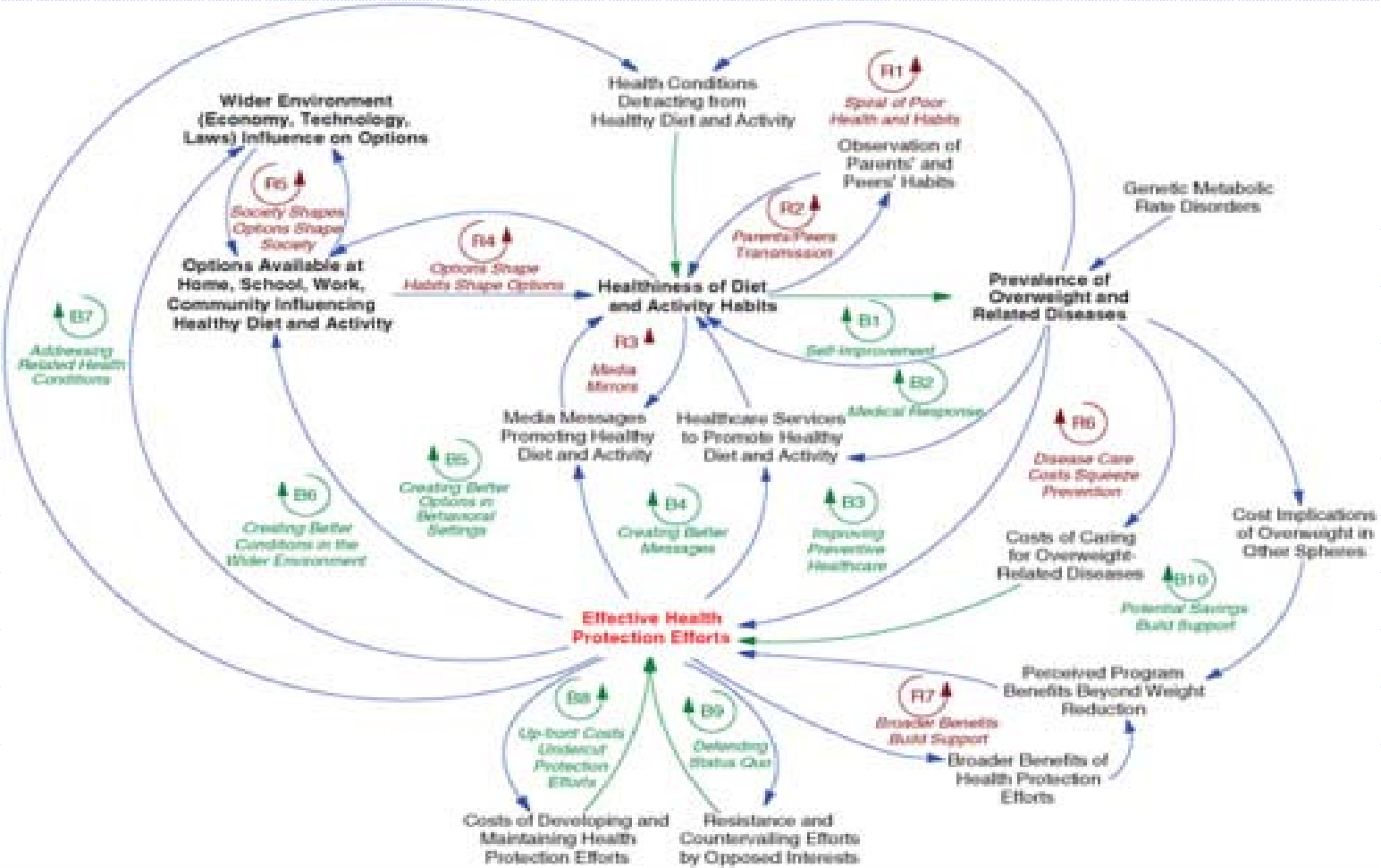
Empirical Application: Experiments/Quasi-Experiments –: Program A causes impact B if B would not have occurred in the absence of A (the counterfactual).

In other words: Program A causes impact B, if there is a *difference* between the value of B in Treatment and Control Groups (Average Treatment Effect on the Treated, ATT).



2. Two Approaches to Causation

Example of a Causal Map of Health Protection Efforts



2. Two Approaches to Causation

Why Talk about Causation?

Foundational issues get lost in applied debates but they matter:

1. Forestall Erroneous Claims

"Any attempt at drawing a causal inference … requires answering essentially counterfactual questions: How would individuals who participated in a program have fared in the absence of the program?: (Duflo et al., 2008)

2. Resituate Debates about Impact Assessment (esp. RCTs)

RCTs as "hard evidence" "gold standard", (Banerjee, 2007), etc. etc.

3. Methodological Differences

Outcomes vs. Mechanisms

- Difference-Makers (Experiments/quasi-Experiments) focus on causal effects.
- **Holland Rubin Framework**: "Others are interested in understanding the details of causal mechanisms. The emphasis here will be on measuring the effects of causes because this seems to be a place where statistics … has contributions to make" (Holland, 1986)
- Focus on magnitude of effect: the **how much** question:

- Producers focus on mechanisms (causal variables, causal tree and explanation of how variables do their causal work).
- Emphasis is on the **how/why** questions.

- RCTs do address causal mechanisms but in a 'thin' way, focus on design features of programs/projects (causal intermediaries), e.g. randomising bookkeeping support services when evaluating credit programs.

3. Methodological Differences

Intersubjective Observables vs. Dialogically Generated Information

Intersubjective Observability as a desired feature of knowledge appears in Empiricism (early Popper), economics (Samuelson, etc.) and also in experimental designs.

Holland/Rubin: fundamental problem of causal inference is, that the same person cannot partake in both treatment and control groups: 'It is impossible to observe the value of Yt (u) and Yc (u) on the same unit, therefore, it is impossible to observe the effect of t on u. The emphasis on the word observe.' (Holland, 1986).

RCTs solve this problem, by creating an observable control group identical in expectation to the treatment group.

Qualitative impact assessment emphasises dialogically generated information from focus groups, semi-structured interviews, etc. to understand locally meaningful information (categories, relationships, practices, etc).

'Double hermeneutic' – interpreting a pre-interpreted world.

3. Methodological Differences

Thick vs. Thin Description

Refers to the level of explanatory detail

- Experiments/Quasi-experiments are thin.
- Focus is on magnitude of causal effect
- Analysis of mechanisms is limited to program design/causal intermediares

- 'Qualitative' is generally thicker
- Understanding of causal mechanisms extends to social relationships, behavioural motivations, micro or meso processes driving outcomes, etc.
- Attention also directed to choosing locally meaningful impact indicators.

4. Mixed Methods in Impact Assessment

Example 1: Combining Difference-Makers (Outcomes) and Producers (Mechanisms)

IFPRI Evaluation of Risk Mitigation Program in Turkey (Adato (2008))
Impact assessment of CCT program (Cash conditional on school enrolment)

Regression Discontinuity Design (RDD) drawing on Household Survey (HHS) data combined with ethnographic work in 6 localities

- RDD found significant effect on enrolment (up 10%) but levels remained very low for girls (40%)

Ethnographic work sought to explain the reasons why

- Low value placed on education in context of high value according traditional roles as mothers and wives
- Perceived threat to family honour/reputation associated with schooling – "girls have only their honour and it is my duty to prevent any bad words about that … [with schooling] they will look at them in a bad way"

Points: i) thicker description of processes/mechanisms driving outcomes combined with estimates of outcomes (Producers + Difference Makers); ii) implies CCT type programs are limited where constraints are not monetary

4. Mixed Methods in Impact Assessment

Example 2. "Difference-Makers": Combining Intersubjective Observables (Quasi-Experiments) and Dialogical Information (Thought Experiments)

Impact Assessment of Hunger Eradication and Poverty Reduction Program in Vietnam (Shaffer, 2012, 2013)

- Impact assessment of Health Fee Exemption Component (free health care to poor households or communes).
- Combined use of Propensity Score Matching (PSM) and 'Counterfactual Thought Experiment' (CTE)'
- Impact measure was health care utilisation was defined as the percentage of persons who used health care facilities over the past 12 months
- PSM drawing on national HHS data found no significant impact (next slide)

- CTE in a 'qualitative survey' asked respondents whether they still would have sought medical attention when they were ill if they had not received the health fee exemption or reduction.
- Around 95% of respondents would have sought medical care when they were ill even if they had not benefited from program (accounting for sampling error and 'don't knows (next slide).

Points: i) enhances validity of results; ii) same concept of causation (counterfactual dependence) but different methods of causal inference;

4. Mixed Methods in Impact Assessment

Example 2. Combining Intersubjective Observables (Quasi-Experiments) and Dialogical Information (Thought Experiments) (Cont.)

Impact Assessment of Hunger Eradication and Poverty Reduction Program in Vietnam (Shaffer, 2012, 2013)

Table 3 Propensity Score Matching: Impact of Health Fee Exemption/Reduction on Utilisation of Healthcare

	Mean Difference	Standard Error[a]	95% Conf. Interval
Nearest Match	-0.09	0.06	-0.23 - 0.02
Nearest Three Matches	-0.08	0.05	-0.20 - 0.01
Nearest Five Matches	-0.08	0.05	-0.18 - 0.01

[a]Standard errors were bootstrapped with 100 replications
Data source: Vietnam Household Living Standards Survey, 2002

Table 2 Use of Medical Care in the Absence of HEPR (Population Proportions, Standard Errors in Parentheses)[a]

	1 Yes	2 No	Total
Total Vietnam	91.8	7.3	100
	(0.40)	(0.11)	

[a]Data do not sum to 100 because "Don't Knows" have been removed
Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

4. Mixed Methods in Impact Assessment

Ex. 3. Using Thick Description to Better Understand the Causal System

Study of the Determinants of Welfare Changes, Kagera Region, Tanzania (de Weerdt, 2010)

- Econometric analysis of panel data (1994/2004) combined with life histories
- Econometric analysis estimated 2004 poverty status based on 1994 asset holdings and compared it to actual outcomes
- Significant disparities resulted – only 50% of expected exits happened.
- Life histories used to explain why
- One key finding – the interaction between remoteness and initial conditions, i.e. land and other assets, (and not their individual effects) was critical
- That is, non-remote village could overcome low initial condition due to trade (i) employment for labours (loading trucks and carrying good), ii) development of business relationships with traders leads to trading 'start-ups'; iii) influx of money increases demand for brewing/distilling; iv) new ideas & innovation +access to trade networks)
- Accordingly, the causal structure in the original model was incorrectly specified.
- A new interact variable of remoteness and initial conditions was introduced and they model was found to be a better fit for the data.

5. A Case for Mixed Methods in Impact Assessment

1. The World is Complex and Causal Inference is Difficult

- complex causation (many different variables are involved);
- multiple causation (these variables are affecting each other in many ways);
- simultaneous causation (these effects between variables are occurring at the same time).
- spurious causation (an unknown 'third' variable may be the true cause)
- intentionality/agency (in the social sciences, causal variables/effects involve human agents)

Strong case that no one concept of causation or model of causal inference is sufficient to provide a full account across all causal systems which entails:

1. Identifying causal **variables**
2. Specifying their interrelationship (Causal **Tree**)
3. Assigning causal **weights** to the variables
4. Identifying underlying causal **mechanisms** (how/why causal effect is brought about)

5. A Case for Mixed Methods in Impact Assessment

2. Empirical Adjudication/Determining Validity is Difficult

"There are no pure facts, only facts as couched in one conceptual scheme or another. There are no pure observations, but rather observations couched in a theory-laden vocabulary. Theories bring with them their own empirical criteria."
(Little, 1998)

Validity is strengthened if different approaches arrive at the same results (triangulation).

4. A Case for Mixed Methods in Impact Assessment

3. Mixed Methods (can) Add Value

Can provide better, richer analyses, e.g.:

- Integrating Outcomes and Processes (How much and How/why)
- Aiding in Empirical Adjudication/Determining Validity
- Facilitating Understanding of the Causal System

But, requires an appreciation of why results may differ, such as different:

- Interpretation of words (household, work, sick, etc.)
- Time periods, population coverage, dimensions of well-being, etc.

And, depends on the endgame:

Research question should determine methods, not vice versa