Integrating evidence to tell the evaluation story

Greg Mason
Presentation to Canadian Institutes of Health Research
November 8, 9, 2012

Day 2 – Morning
Editing, assessing, interpreting

Goals of the workshop

- To present an intermediate and critical review of the main qualitative and quantitative lines of evidence currently used in evaluations
- To align these lines of evidence with the nature of the questions posed
- To review how to draw the lines of evidence into a coherent evaluation “story”

Outline of the workshop

Day 1 Morning – Creating the evaluation plot

Day 1 Afternoon – Adding characters

Day 2 Morning – Editing, assessing, interpreting

Day 2 Afternoon –Telling the story

Working to the climax: the methodology of value determination

- **Value-for-money** relies on
  - Assessing program/policy relevance
  - Assessing outcomes
- Assessing program/policy **relevance**
  - Literature review
  - Expert interviews
  - Document review
- Assessing **outcome achievement**
  - Net impact using quasi-experimental design
  - Contribution analysis

Literature review

- **Goal**
  - Chart origins and antecedents
  - Practice/experience in other “cognate” jurisdictions
- **Steps in document identification**
  - Keyword search of databases using Boolean processes
  - Common databases include
    - Google scholar
    - PubMed
    - Scopus, etc.
- **Citation managers**
  - Zotero
  - Qippa
  - Thomson Reuters

Steps in literature review

- **Create keywords from…**
  - Evaluation matrix
  - Foundation documents (TB subs, etc.)
  - Senior consultations
- **Scan literature** (last decade is usually sufficient)
- **Triage documents** into highly relevant, related, and unrelated
- Consider using a **citation manager** to catalogue and annotate the discovered articles
- **Continuously scan** materials of high relevance
- **Revise keywords – re-search**
- **Actively hypothesize,** relating the new materials to the evaluation questions and updating the search terms

Warning – it is easy to wander.
Stay focussed on matrix and issues.

Expert interviews

- For science-based programs, these can be invaluable
- Experts will often have made contributions to the literature
- Invaluable to interpreting trends and debates, and offering insight into future direction
- May offer an unbalanced view of policy and programming (strong views can influence evaluator)
- Also may serve a role as methodology advisors

Case study – BSE I, II (Health Canada)

Focus on examining the science-base to the BSE I and BSE II Initiatives to answer

- Q1: Is there a continued need for the BSE I and BSE II Initiatives?
- Q6e: To what extent are there internationally harmonized standards and regulations addressing BSE/TSE and related risks?
- Q6i: To what extent are food and health products safer?
- Q10: Are there alternate ways to deliver the BSE I and BSE II Initiatives to achieve similar results at lower cost?

Search terms for literature review	
Bovine spongiform encephalopathy	BSE risk management
Prion diseases	TSE risk management
Response to BSE	Cost-effectiveness BSE
Response to TSE	Cost-effectiveness TSE
Response to mad cow disease	Cost-effectiveness mad cow disease
Government response to BSE	International BSE standards regulations
Government response to TSE	International TSE standards regulations
Government response to mad cow disease	United States FDA BSE standards regulations
Canadian government response to BSE	United States FDA TSE standards regulations
Canadian government response to TSE	European Union BSE standards regulations
Canadian government response to mad cow disease	European Union TSE standards regulations
	Japan BSE standards regulations
International response to BSE	Japan TSE standards regulations
International response to TSE	Australia BSE standards regulations
International response to mad cow disease	Australia TSE standards regulations
BSE risk assessment	Canada
TSE risk assessment	Health Canada
	Public policy

Steps in document review

- Typical documents may include:
  - Treasury Board submissions and Memoranda to Cabinet
  - action plans, strategic plans, work plans, and operational plans
  - documentation describing governance/management structure, such as Terms of Reference and Memoranda of Understanding
  - communications plans/communications
  - meeting agendas and minutes
  - previous evaluations
  - documents produced as a result of program funding, such as risk assessments, research reports, and guidance documents
- Analysis will link documents to evaluation questions and indicators.
- Use citation managers and NVIVIO to organize and code information

The goal is to respond to the issues from the matrix designated for the document

Relevance, rationale, implementation, delivery

Assessing outcome achievement

The core of evaluation is the **counterfactual**

"We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed." David Hume *An Enquiry concerning Human Understanding* 1748, Section VII.

Where c and e are two distinct actual events, e causally depends on c if and only if, if c were not to occur e would not occur.

Lewis D.(2004): "Causation as Influence"

Counterfactual – the state of affairs that would have occurred without the program

Scientific truth always goes through three stages. First, people say it conflicts with the Bible; next they say it has been discovered before; and lastly they say that they always believed it

Louis Agassiz, Swiss naturalist

We do not now a truth without knowing its cause

Aristotle, Nicomachean Ethics

Development of Western science is based on two great achievements: the invention of the formal logical system (Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance)

Albert Einstein

Preliminary causal glossary

- **Independent (exogenous, cause) variables** – are the direct policy/program interventions and socio-economic control
- **Dependent (endogenous, effect) variables** – represent the outcomes
- **Intervention variables** – are a special class of independent variables that represent policy/programming, often as a discrete (dummy) variable marking the boundary between the program and counterfactual
- **Gross impact** – observed change in the outcome (s)
- **Net impact** – portion of gross impact attributable to the program intervention
- **Experiment** – the purposeful manipulation of independent and intervention variables to observe the change in outcomes.
- **Quasi-experiment** – the replication of manipulation within the context of a statistical model.

Program Theory and Logic Models

- **Theory** explains the intervention and what outcomes are expected
- **Logic model** – two perspectives
  - explains the intervention (***causal logic***)
  - explains the organization of the intervention and how it integrates with broader objectives of government ***(logistical logic)***
- **Performance measurement**

Causal Logic

- **Verbal** – explains the intervention and how it interacts with external events
- ***Graphical*** – presents a "picture" of the program
- ***Abstract*** (mathematical) – formalism that is most useful when quantitative data are available.

Cause and effect

Necessary causes:

- For ***X*** to be a necessary cause of ***Y***, then if ***Y*** occurs, ***X*** must also occur. The fact that ***X*** occurs does not imply that ***Y*** will occur.

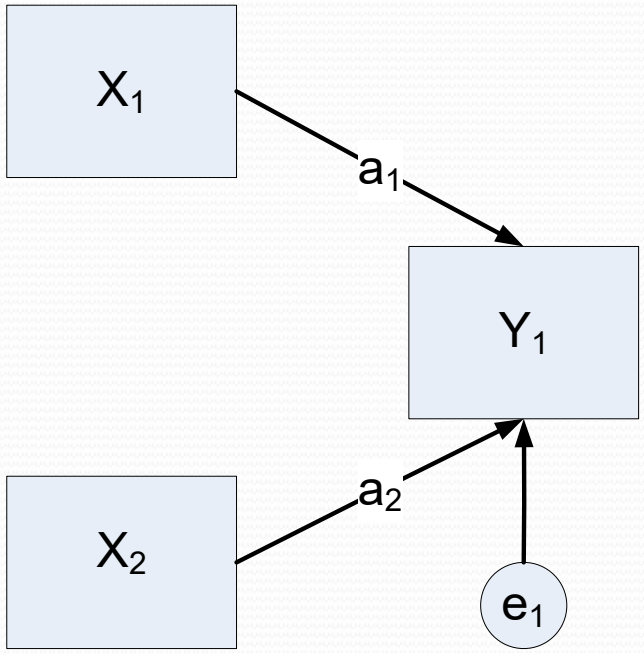
Sufficient causes:

- For ***X*** to be a sufficient cause of ***Y***, then the presence of ***X*** always implies that ***Y*** will occur. The fact that ***Y*** occurs does not imply that ***X*** has occurred since another variable, ***Z***, could be the cause.

Contributory causes:

- A cause ***X*** may *contribute* to the occurrence of ***Y,*** if ***X*** occurs before ***Y*** and varying ***X*** varies ***Y***.

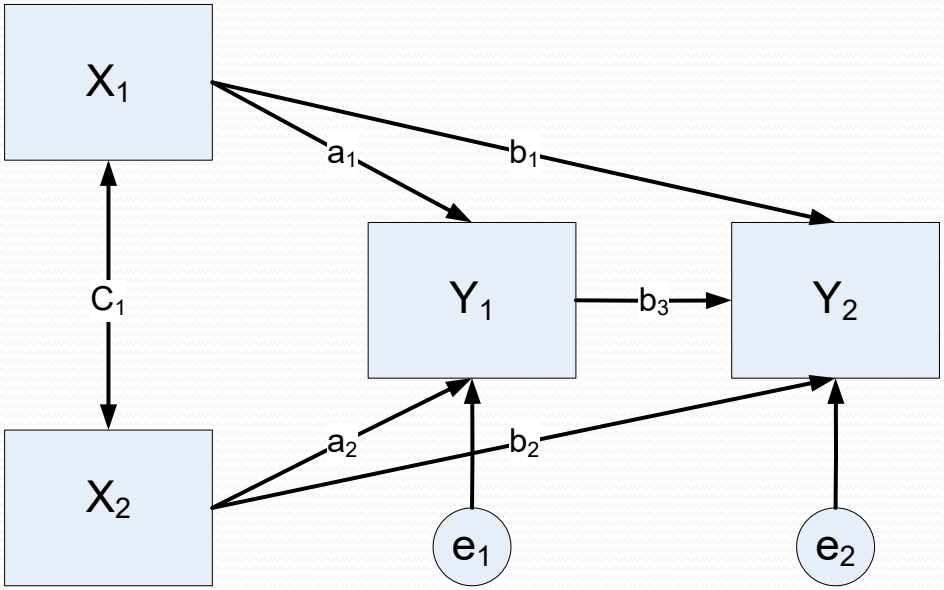
Causal Analysis I



- X_1, X_2 are independent (causal) variables also known as exogenous variables.
- Y_1 is a dependent (effect) or endogenous variables.
- e_1 is an error term, reflecting measurement imprecision, poor model design, failure to include all the relevant variables, external factors…

$$Y_1 = a_0 + a_1X_1 + a_2X_2 + e_1$$

Causal Analysis II



X_1, X_2 are independent (causal) variables also known as exogenous variables.

Y_1, Y_2 are dependent (effect) or endogenous variables.

e_1 and e_2 as above

$$Y_1 = a_0 + a_1X_1 + a_2X_2 + e_1$$

$$Y_2 = b_0 + b_1X_1 + b_2X_2 + b_3Y_1 + e_2$$

$$X_1 = c_1X_2$$

Sociological Path Analysis

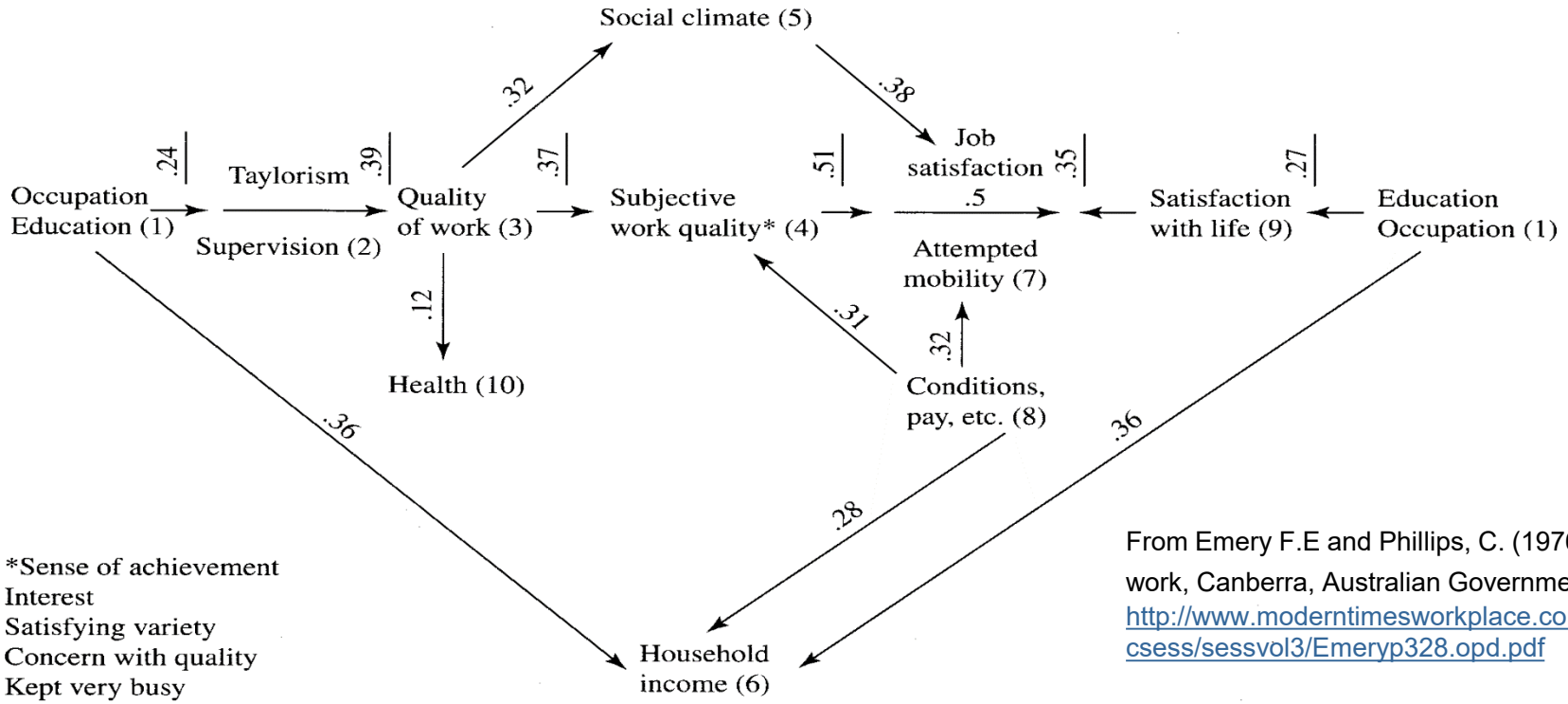


Figure 1. Graph derived from Table 1.

From Emery F.E and Phillips, C. (1976) Living at work, Canberra, Australian Government, http://www.moderntimesworkplace.com/archives/ericsess/sessvol3/Emeryp328.opd.pdf

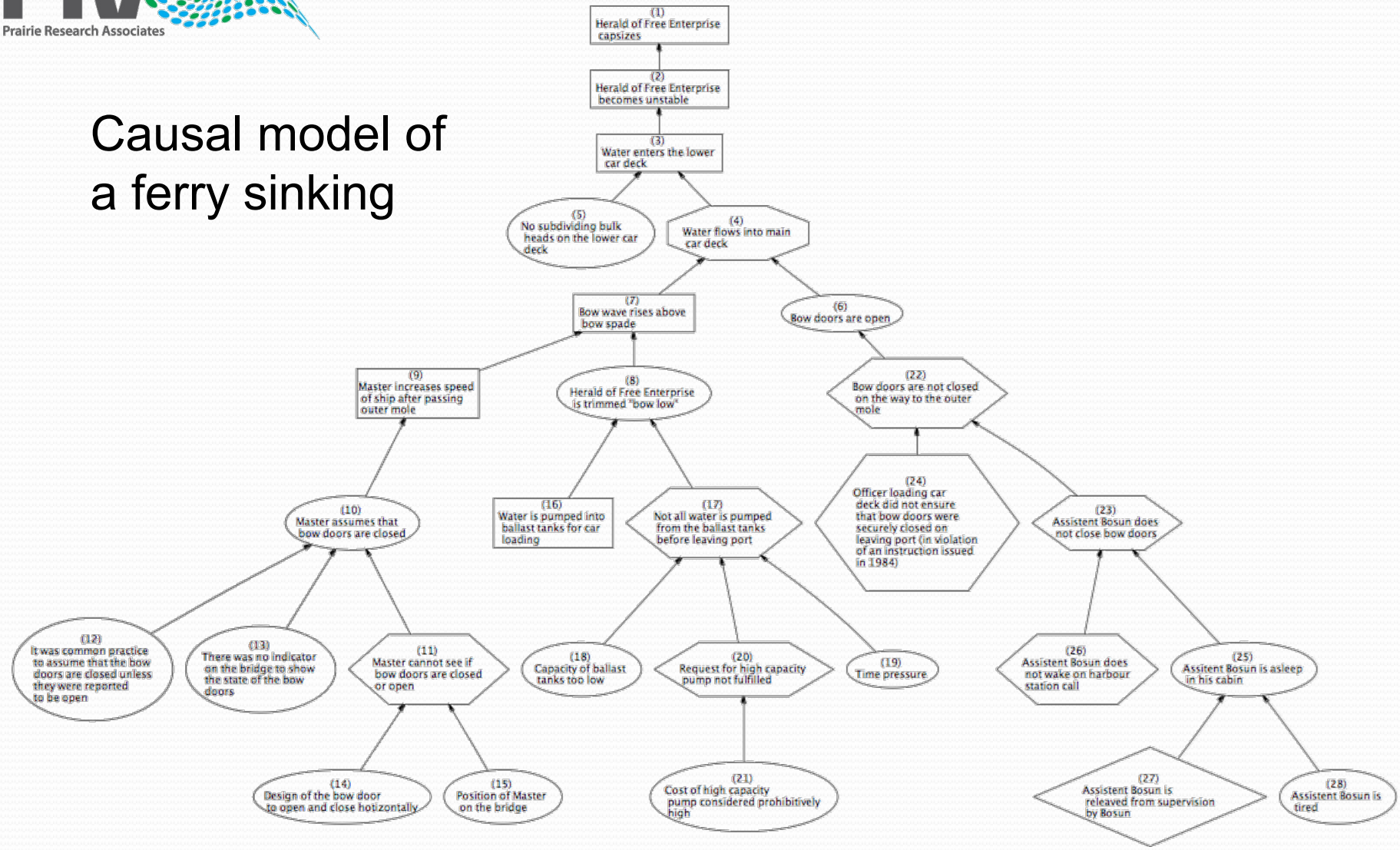
The model shows three contributory causes to changes in household income

..\..\Users\gmason.PRAINC\Documents\Pavtube\youtube_converter\Herald of Free Enterprise Disaster part 1 - YouTube.mp4

..\..\Users\gmason.PRAINC\Documents\Pavtube\youtube_converter\Herald of Free Enterprise Disaster part 2 - YouTube.mp4

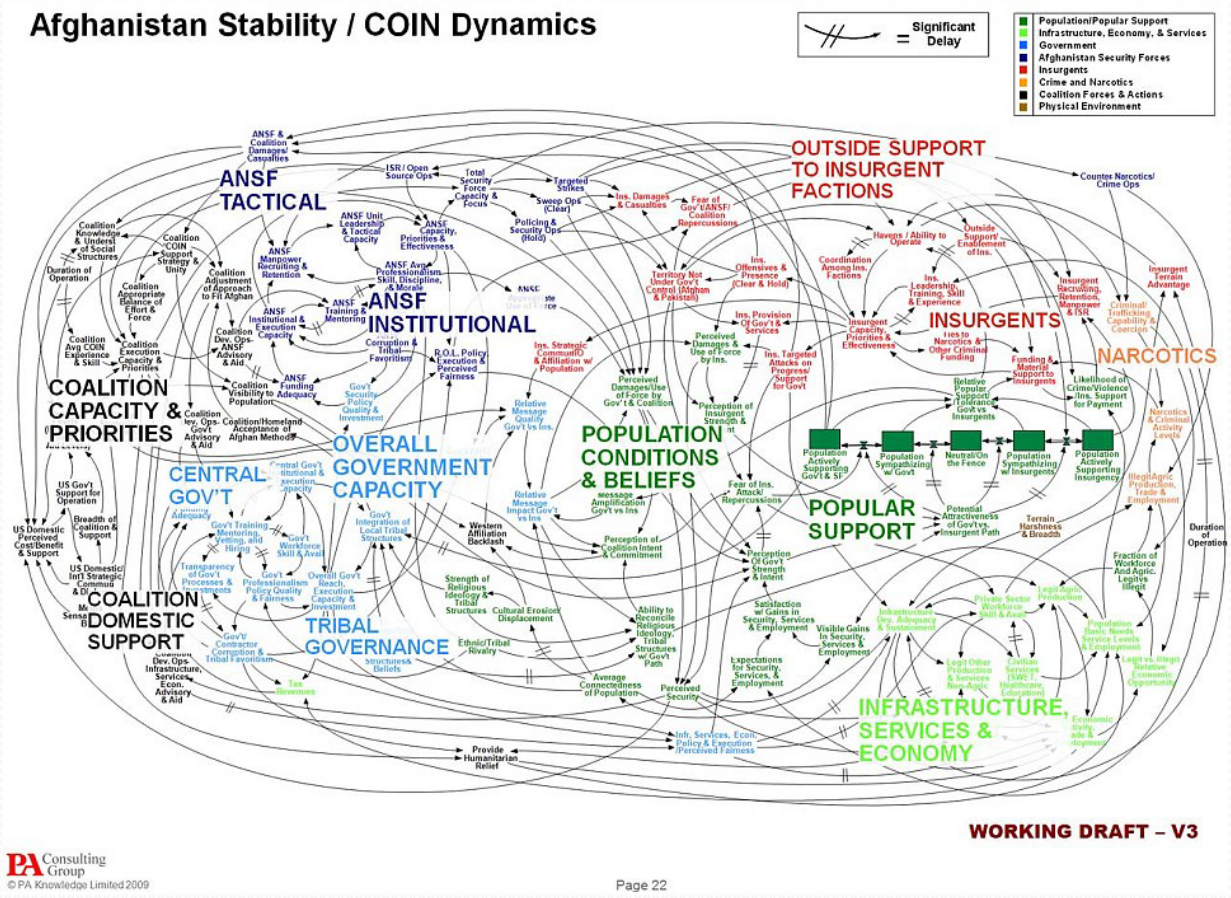
Prairie Research Associates

Causal model of a ferry sinking



Herald of Free Enterprise sinking – causal analysis

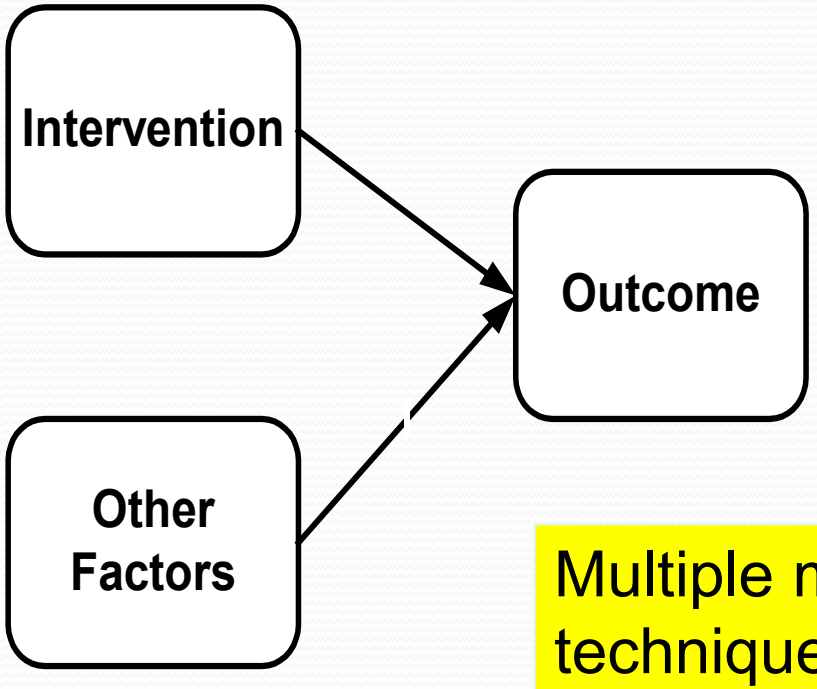
The mother of all causal diagrams



A PowerPoint diagram that portrays the complexity of American strategy in Afghanistan has succeeded.

http://www.nytimes.com/2010/04/27/world/27powerpoint.html?src=me&ref=general

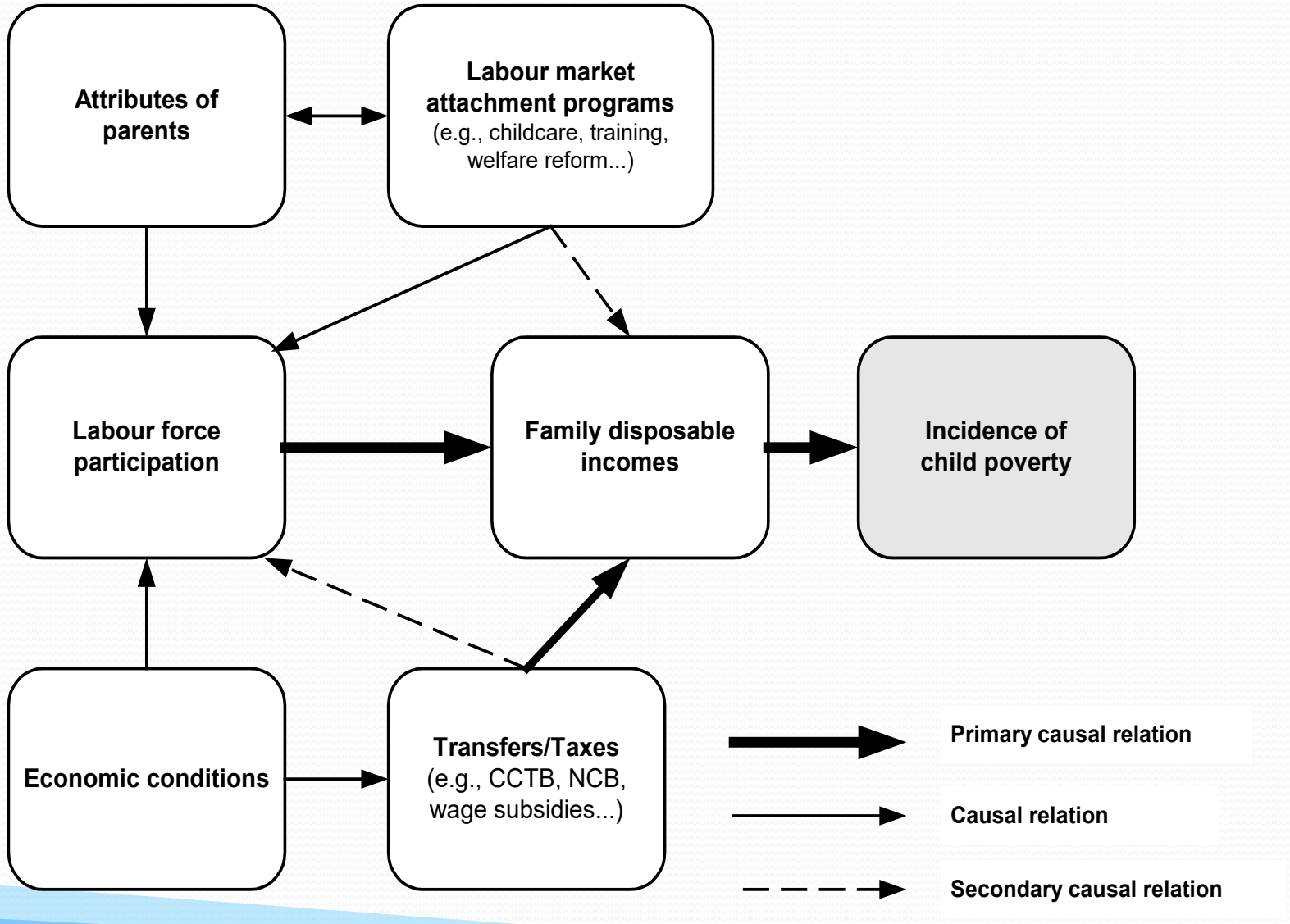
Returning to causal logic models



The causal logic model clarifies the theory of how interventions produce outcomes.

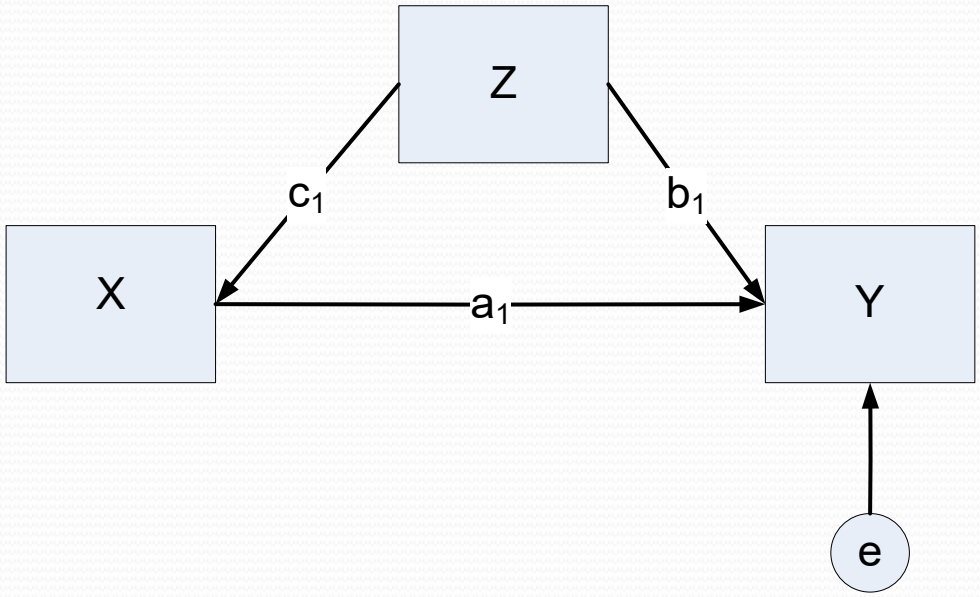
Multiple methods and experimental techniques establish the relative importance of causes of changes in outcomes

Graphical logic for the National Child Benefit



Causal Analysis 3

A confounding variable is a is a variable that correlates positively or negatively with both the dependent and independent variable



$Y = a_0 + a_1X + a_2Z + e$

$Y = b_0 + b_1Z + e$

$X = c_0 + c_1Z$

The problem is that the relationship of interest is X → Y, the confounding variable Z gets in the way.

- effect of X on Y
- effect of Z on Y
- effect of Z on X on Y

Advantages and disadvantages of causal logic models

Advantages

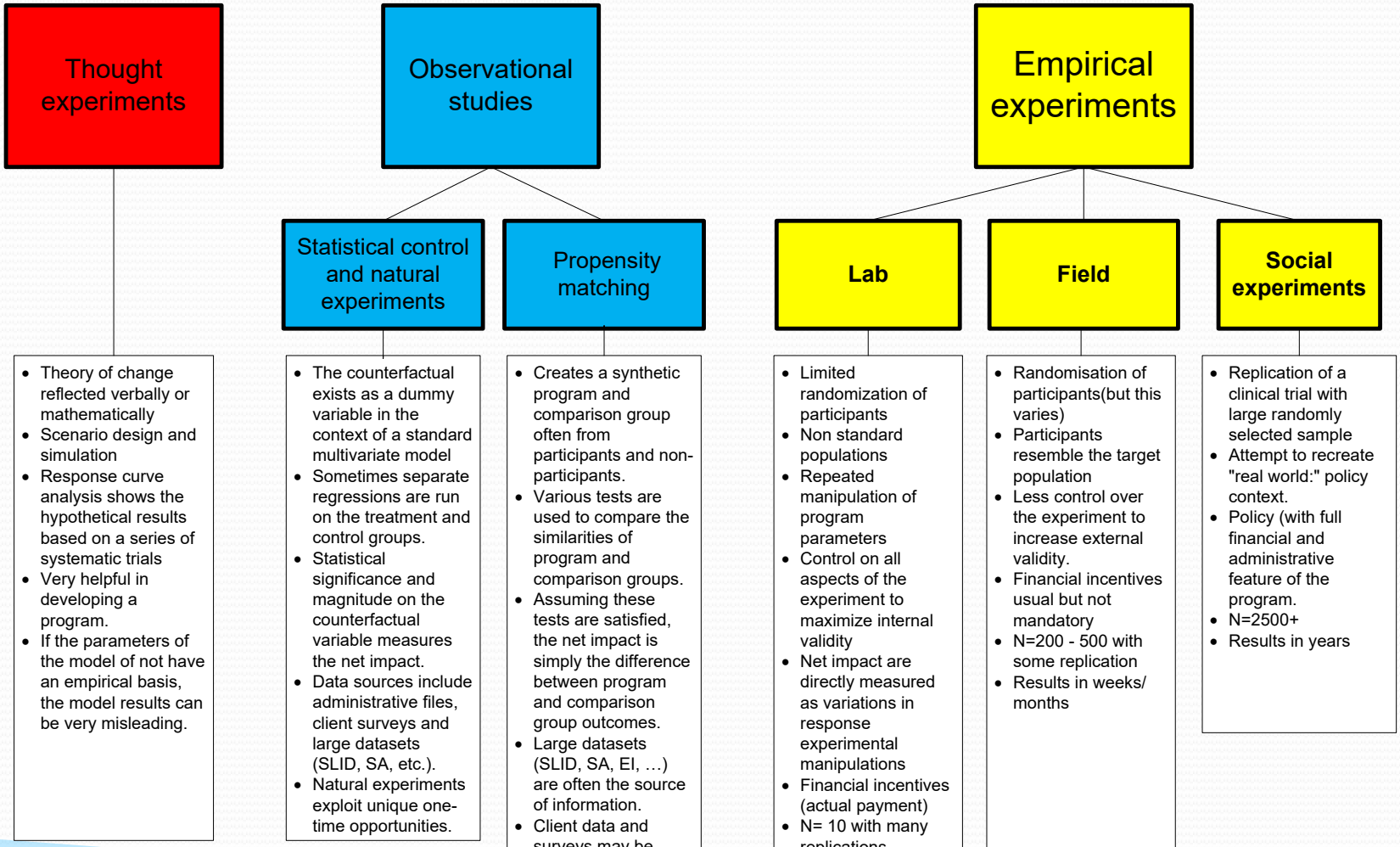
- reveals inter-relationships among program elements
- identifies confounding factors that reduce program outcomes
- identifies the main causal channels
- supports active hypothesizing about magnitudes of effects

Disadvantages

- over-complication can impede understanding
- abstract representations can confine communication
- does not reveal resource use, reach or support other "oversight" requirements
- does not support the discovery of other factors
- does not control confounding

Part 2 – Applied Attribution Analysis

Causal Framework for Policy Design



Three potential models for evaluating policy

1. Randomized control (RC) (social experiments)
2. Statistical control (difference-in- difference, regression discontinuity)
3. Quasi-experimental methods (Propensity score matching)

Random experiments

The classic experiment is the random, double-blind experiment (RDE):

- subjects are selected randomly into a treatment and control group
- each subject receives a code
- an independent third party assigns codes randomly to treatment and control group members.
- the treatment is not identifiable (i.e., the real and fake pill are identical.
- those administering the treatments and placebo have no knowledge of what subjects receive.

Randomization and statistical equivalence

- Randomization into a treatment and control group creates two groups that are statistically equivalent:
  - For any statistic (mean, variance, etc.) the two groups will return results that are the same (within bounds of statistical significance)
  - The test of statistical equivalence applies to observable and unobservable attributes.

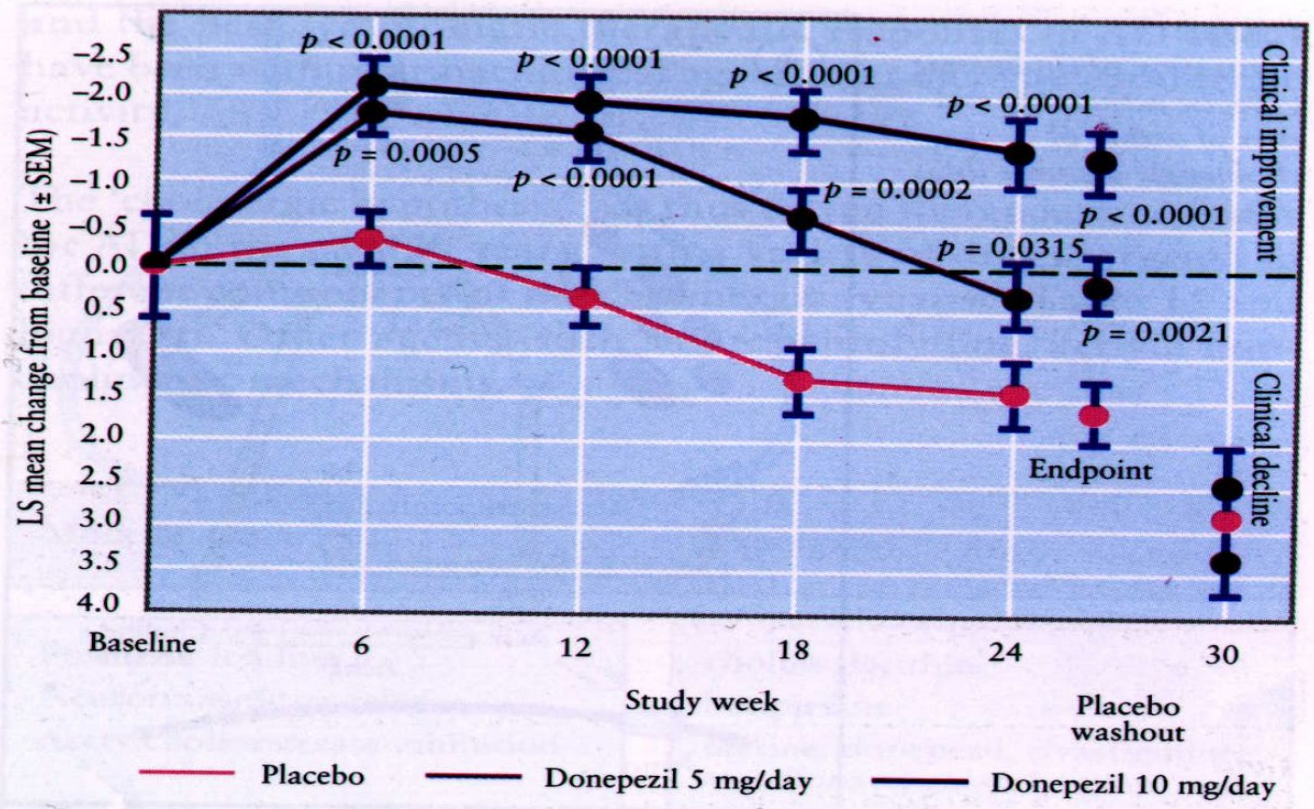


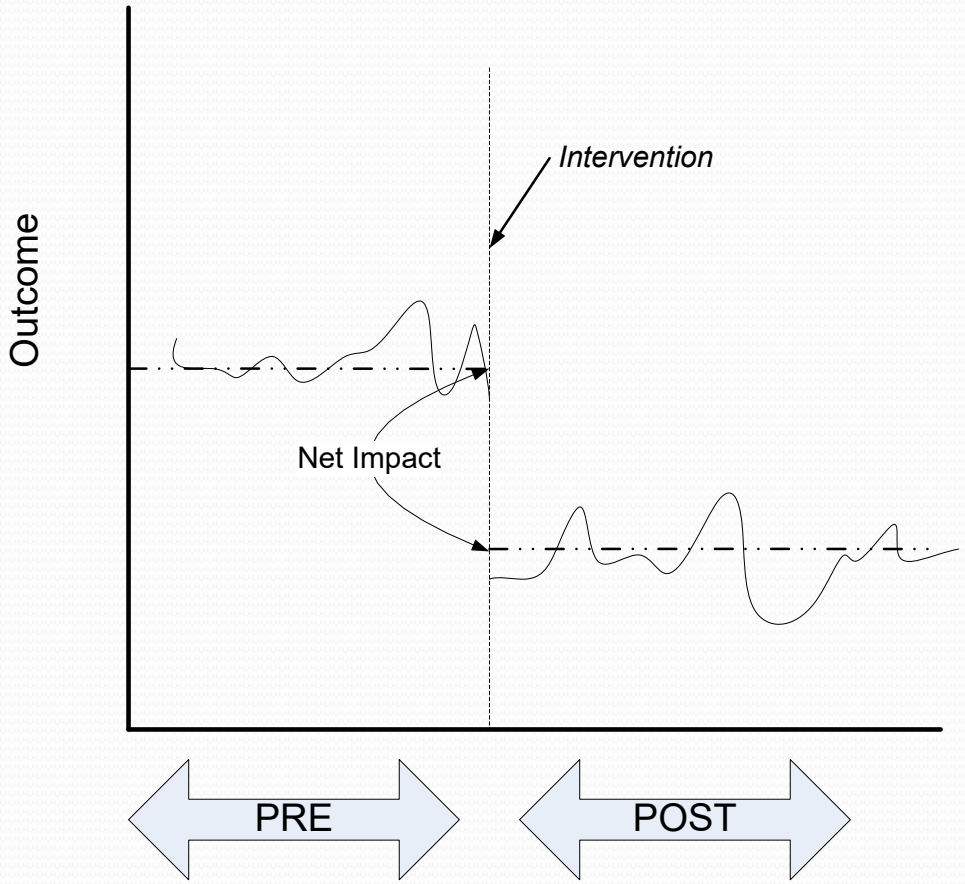
Figure 10
Illustration of therapeutic response to donepezil in an international double-blind, placebo-controlled trial, using the ADAS-cog at regular intervals followed by a 6 week placebo washout period (Gauthier et al 1998).

Limits of Randomized Designs

In social science, randomized double blind experiments are often not feasible:

- human subjects are unreliable (they move, die or otherwise fail to participate in the full experiment).
- many see the administration of a placebo as withholding a treatment.
- social policy cannot be masked (creating a placebo is difficult).

The pre-post design



This model is in wide use. Common examples are seat-belt laws, introduction of legislation (minimum wage). The outcome is critical.

Two common problems are

- Decay
- Identifying the intervention (some interventions have a long implementation)

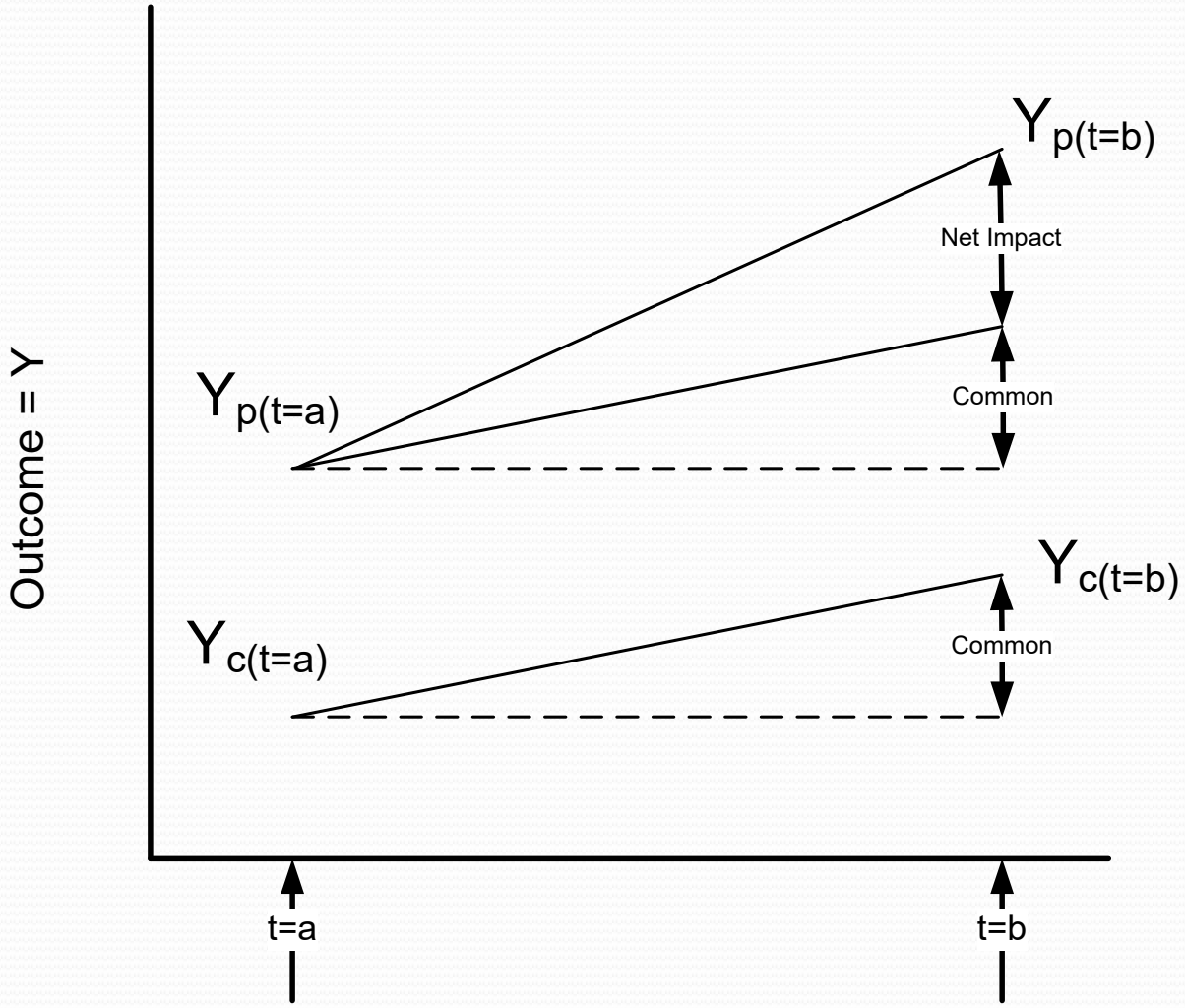
Natural experiments

- Create a “split” in the sample where treated and untreated are classified by a variable that is not related to the treatment.
- This split occurs “naturally” where the program change occurs in one area/jurisdiction, not in others that are “closely similar.”
- Difference-in-differences (DID) methods are a common evaluation framework.

Minimum Wages – case study

The conventional economic wisdom is that an increase in the minimum wage will increase unemployment and reduce incomes (increase poverty). A natural experiment tested this by comparing the outcomes of a minimum wage increase on the employment and wages of teenagers working in fast food restaurants in adjacent areas (New Jersey and Philadelphia) after one state increased the minimum wage. The result was an unchanged level of employment.

Difference in differences



Moving Up A Gear Can a free bike help a girl's education in northern India - YouTube#!.mp4

Matching

In social experiments, participants differ from non-participants because:

- Failure to hear of program
- Constraints on participation or completion
- Selection by staff

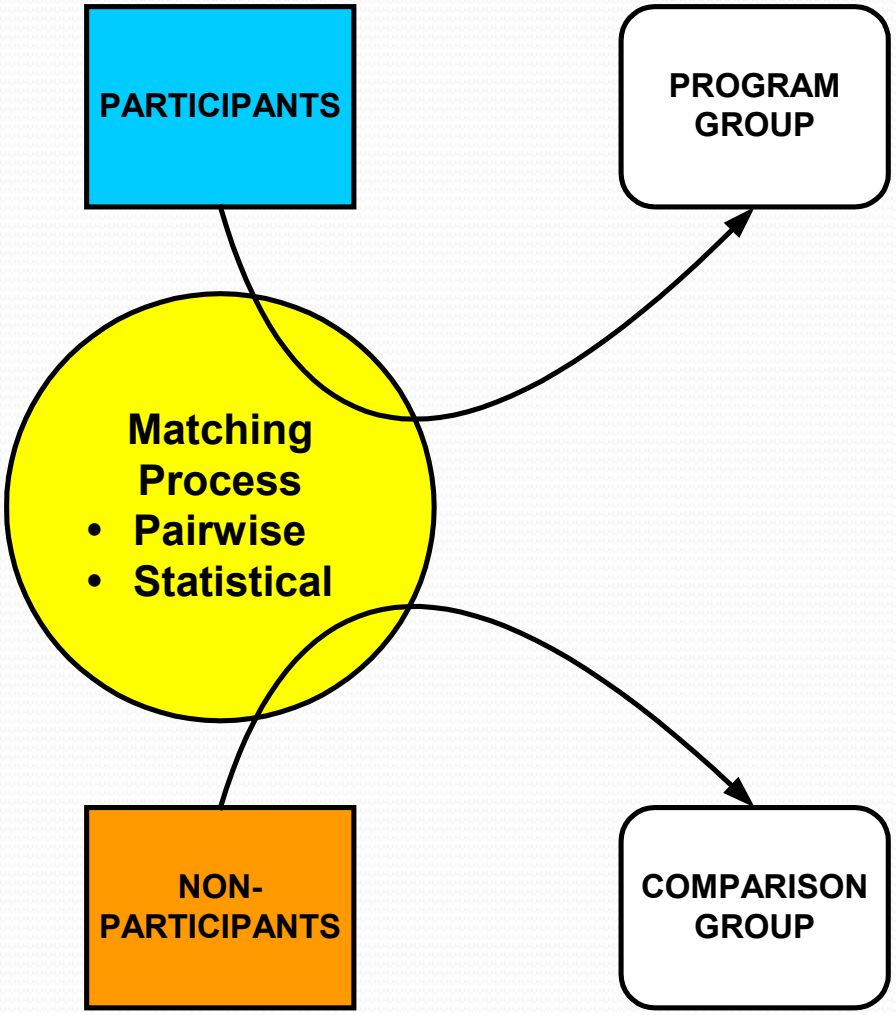
Creating a matched sample of participants and non-participants can be accomplished via

- Pair-wise alignment (exact matching)
- Statistical matching
- Hybrid – exact and statistical

Statistical matching

- Matching is needed because we cannot randomly allocate clients to the program and comparison groups. Program benefits cannot be withheld.
- Logit model provides the estimate of the propensity to participate for participants and non-participants.
- The key idea is that we estimate that propensity to participate is based on observed attributes of the participants and non-participants.
- Participants are assigned a “Y”value of 1 and non-participants are assigned a “Y” value of 0.
- A logistic regression then estimates the propensity to participate.
- Note that even though a non-participant actually did not participate, the model will assign a score between 0 and 1. Typically, non-participants will have lower scores than participants, but there will be an overlap.
- The overlap is termed the *region of common support*.

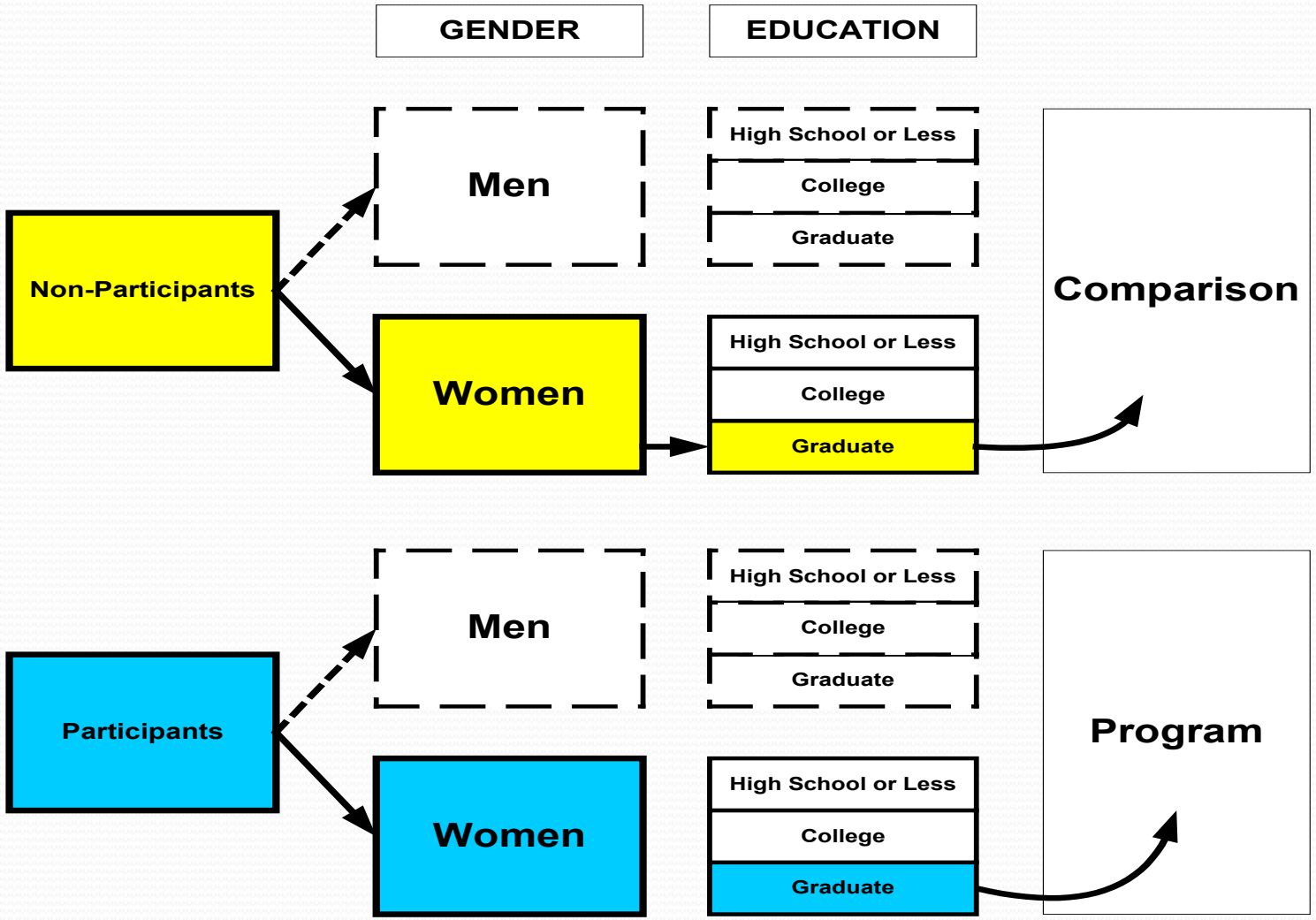
Matching Process



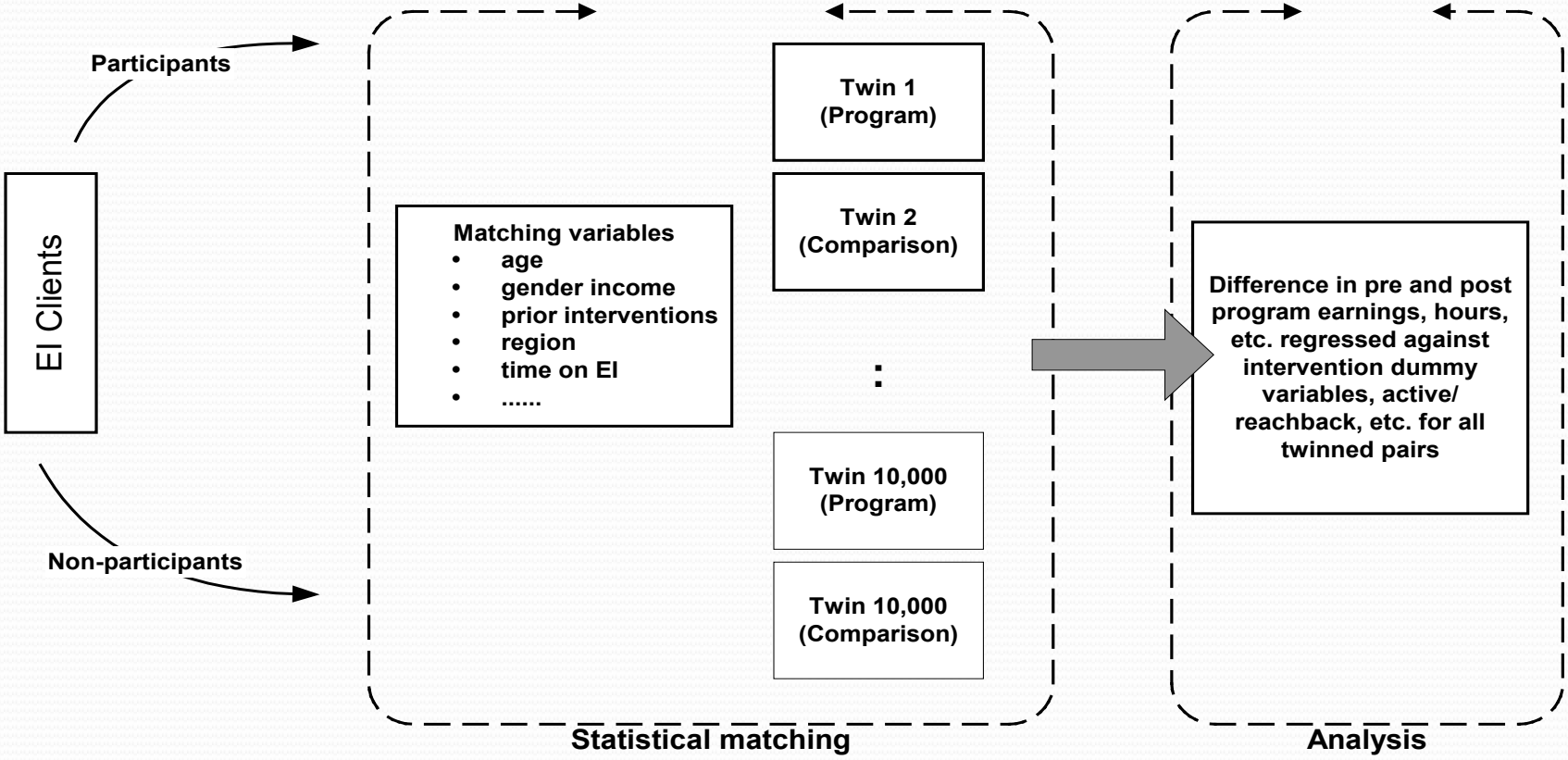
Pair wise matching

- The theory will indicate those attributes that are likely to make a difference in the quasi-experiment.
  - For labour markets, gender, education, and rural-urban location are important
  - For health policy, age, rural-urban, and family history might be important.
- The analyst starts with the first variable and divides the participants and non-participants into two sets.
- Within the sets, the samples are classified with respect to the second variable and so on.

Pair wise Matching



Statistical Matching Applied to the LMDA



Statistical matching and structural modelling