

# Attribution in evaluation: Applied causal inference

Greg Mason

PRA Inc. & University of Manitoba

June 25, 2014

View notes pages have supplementary material and references

# Outline of the module

## **Part 1 – Exploring causal concepts**

- Understanding causality/attribution
- Causal mechanisms
- Causal logic models: verbal, symbolic, and abstract
- Example: National Child Benefit
- Example: Beneficial management practices for agriculture

## **Part 2 – Control, comparison, and contrast**

- Randomized control treatment (RCT)
- Quasi-experimental methods: mimicking the RCT
- Natural experiments
- Statistical control: multivariate methods, regression discontinuity and instrumental variables
- Qualitative approaches to causal inference: contribution analysis, root cause analysis, and survey based methods

## **Part 3 – Case studies**

- Clinical trials for Aricept
- National child benefit
- Flouracial over dose
- Beneficial management programs in agriculture

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 know 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*

## Biases and perspectives

1. An independent social reality exists. The program sphere is a subset of social reality.
2. Individual and collective interpretation of social reality is always partial and incomplete.
3. Individual action has some, but typically very limited influence on states of that reality.
4. Collective action has commensurately greater influence on social reality, but interacts with external trends and random events in the realization of social reality.
5. Purposeful changes (interventions) can affect states of reality, but what to change and how to change states is influenced by individual and shared understanding of that reality.
6. It is possible to represent (interpret) reality as an abstract model comprising measureable and meaningful variables.
7. Explanation equals causal understanding.

# Part 1 – Causal concepts

## The deductive-nomological (D-N) model

- Scientific explanations:
  - Are logically valid and comprise sound argument (deductive)
  - Include one “law”

Since: Laws

L1, L2, .... Ln

Given that: Initial condition

C1, C2, ... Cn

Therefore: E

Explanations

---

- *Example:* For every 1% increase in unemployment, actual GDP will be 2% lower than potential GDP. (*Okun's law*)
- *Example:* Under capitalist development, the return to capital exceeds the return to labour, creating progressive income and wealth inequality. (*Piketty*)

## What are good explanations

- Good explanations feature causal processes (mechanisms) that link initial conditions, changes in certain “state” variables, and the resulting changes in social reality.
- “Regularity” between the change in state and the change in social reality is not enough.
- We need to know why –

### **the mechanism**

- “Why” and “how” come from a theory of change.
- Confirming the mechanism requires *direct* or *indirect* manipulation.

## Program theory and logic models

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

## 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  $Y$ .

## 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** – special class of independent variables that represent policy/programming
- **Discrete (dummy, 0-1) variable** – marks the “boundary” between the program and counterfactual
- **Counterfactual** – the state of affairs that would have occurred without the program/intervention
- **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.

# Causal logic models

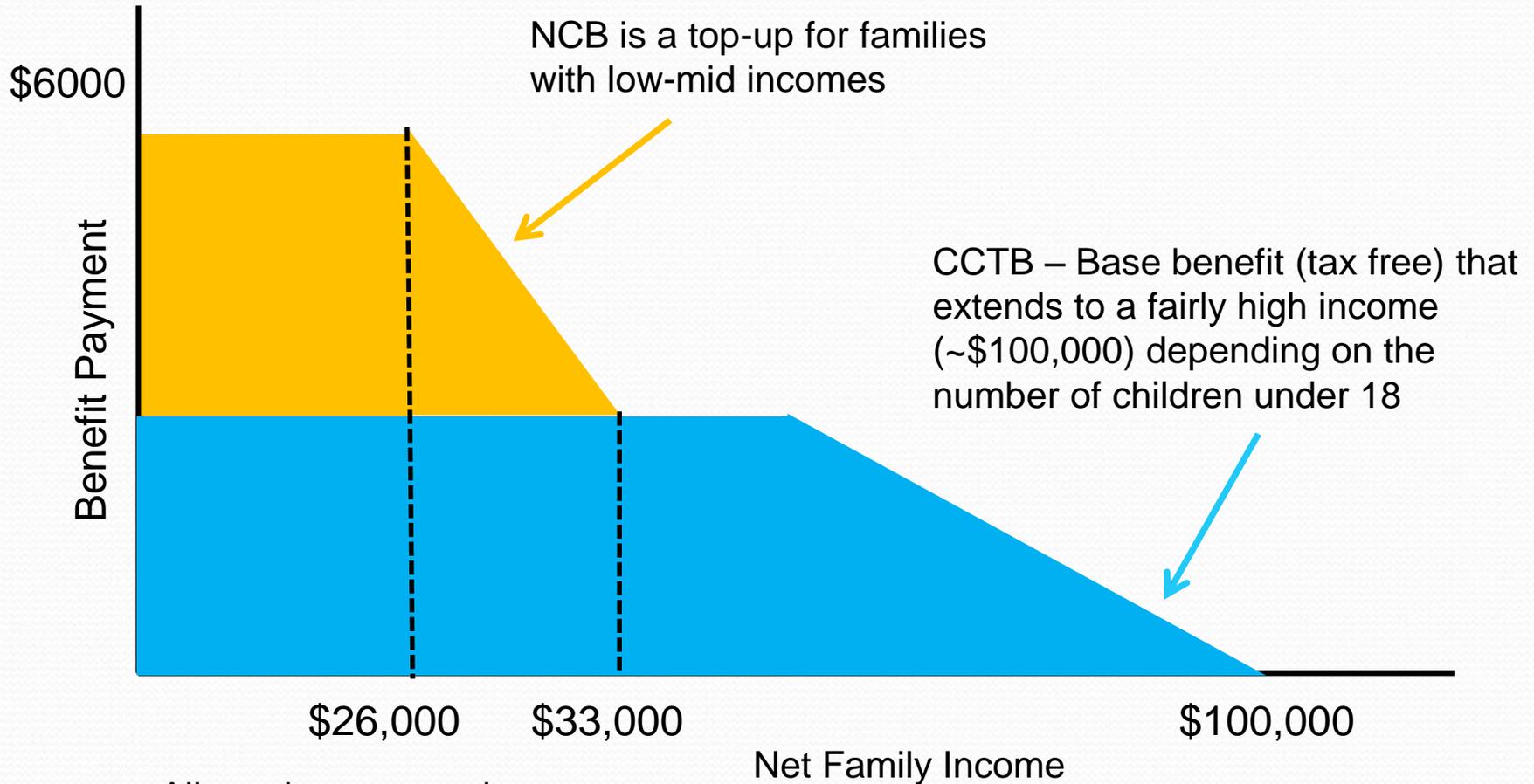
## Verbal models

### **National Child Benefit (NCB)**

The NCB Initiative is a joint initiative of federal and provincial/territorial governments intended to help prevent and reduce the depth of child poverty, as well as promote attachment to the workforce by ensuring that families will always be better off as a result of working.

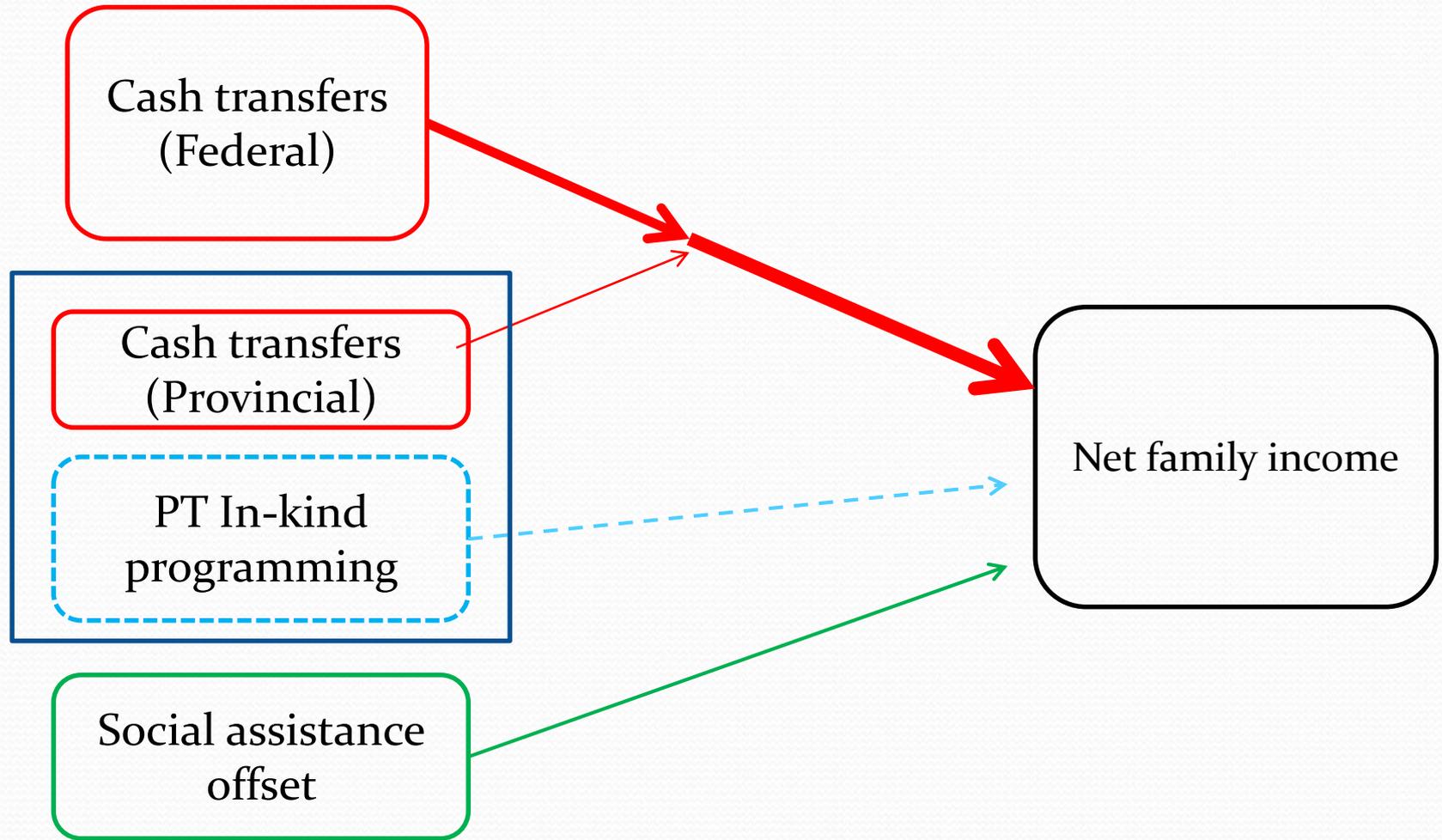
It does this through a cash benefit paid to low-income families with children, a social assistance offset, and various supplementary programs provided by provinces and territories.

# National Child Benefit (two children < 18)



All numbers approximate

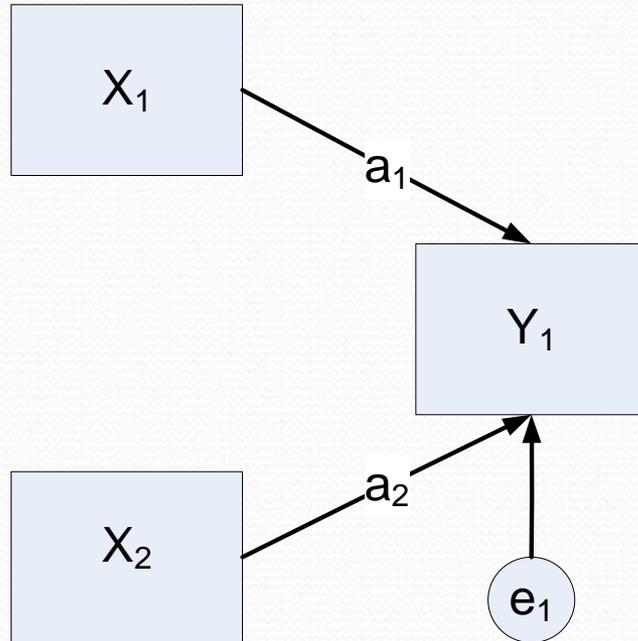
# Mechanism of National Child Benefit



# Theory of change and logic model

- The logic model — more than a static representation of a result chain
- Comprises
  - symbolic explanations
  - text and context
  - possibly an abstract (mathematical model)
- The NCB “logic” of the program reveals the mechanisms whereby the intervention is believed to affect the final result:
  - Reduction in the depth and incidence of child poverty  
.... measured by ....
  - Increase in net family income
- The intervention is complex
  - Direct cash transfers
  - No cash support for those on social assistance
  - In-kind programming to support employment, health, childcare, early childhood development, etc.

# Causal Analysis I



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

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

# Root cause analysis

## Root cause analysis

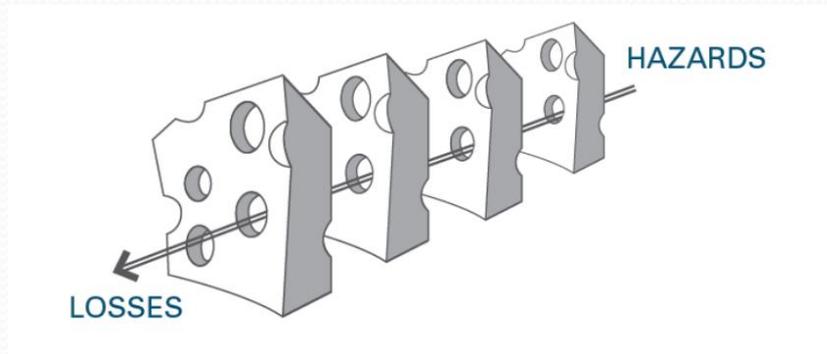
- structured process for reviewing an event (accident), to determine what happened, why it happened, and how to reduce the likelihood of recurrence

### Common to:

- Aircraft crashes and other transportation incidents
- Hospital incidents (medication, surgical mishaps)

## Basics of root cause analysis — Swiss cheese model

1. “Defences (firewalls) are not impermeable and can be penetrated when active failures (unsafe acts) and latent conditions (dormant system conditions) combine to create the opportunity for an incident.”
2. Latent conditions can be identified and corrected
3. Humans make mistakes
4. Key issues in every incident
  - how and why the defences in the system failed
  - look at the system as a whole, rather than just at the actions of individuals.
5. Organizations with low incident rates expect, “own,” and manage failures.



## Methods of root cause analysis

- Collect initial understanding — what and how did the event occur
- Detailed understanding to create a comprehensive set of linked causal factors — expert opinion is essential
- The incident is causally diagrammed with temporal detail (what happened, when, and what is the timing in the steps)
- Link corrective steps to each element within system control

# Sociological path analysis

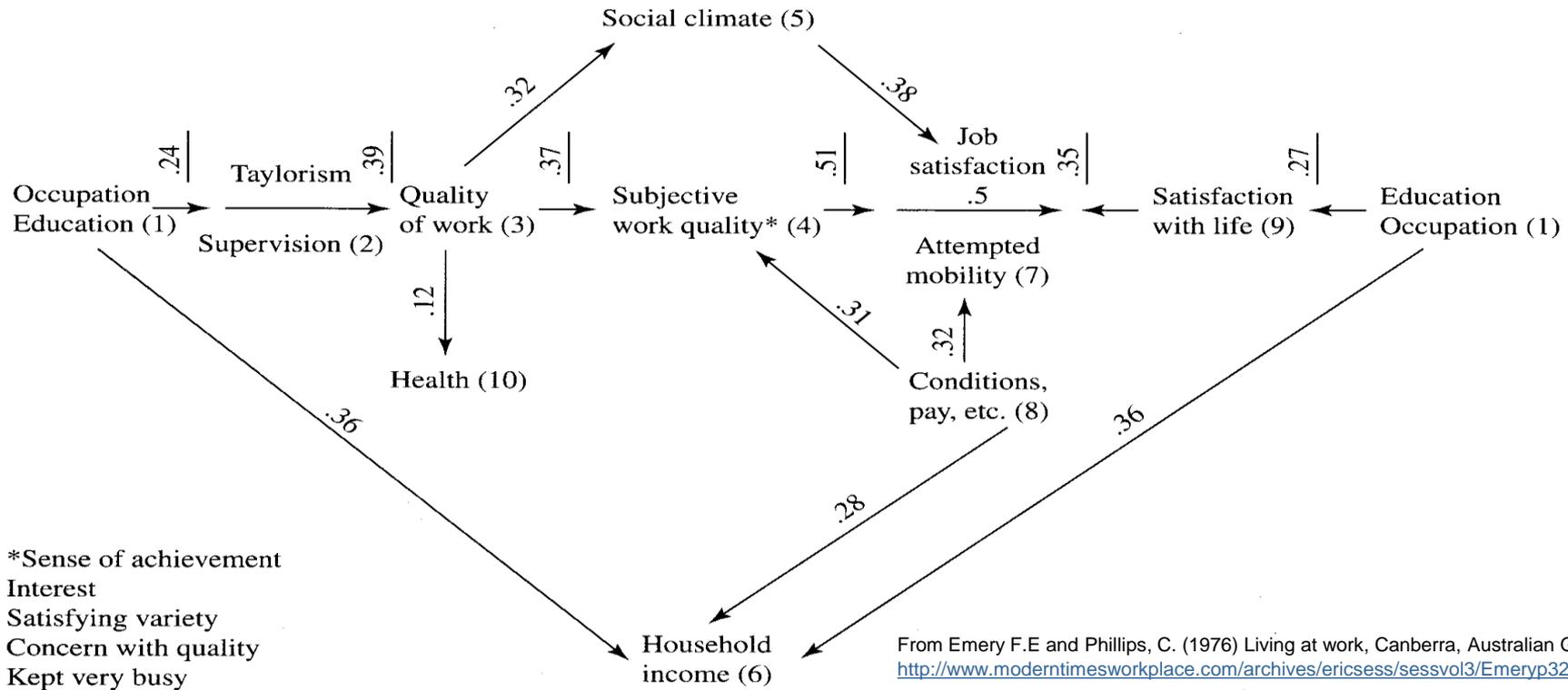
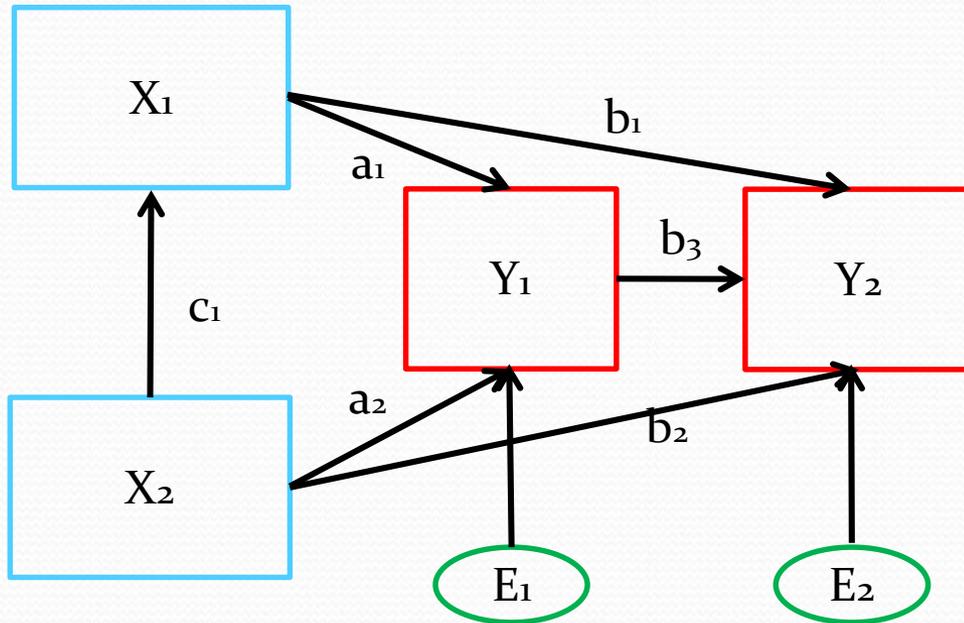


Figure 1. Graph derived from Table 1.

## 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.

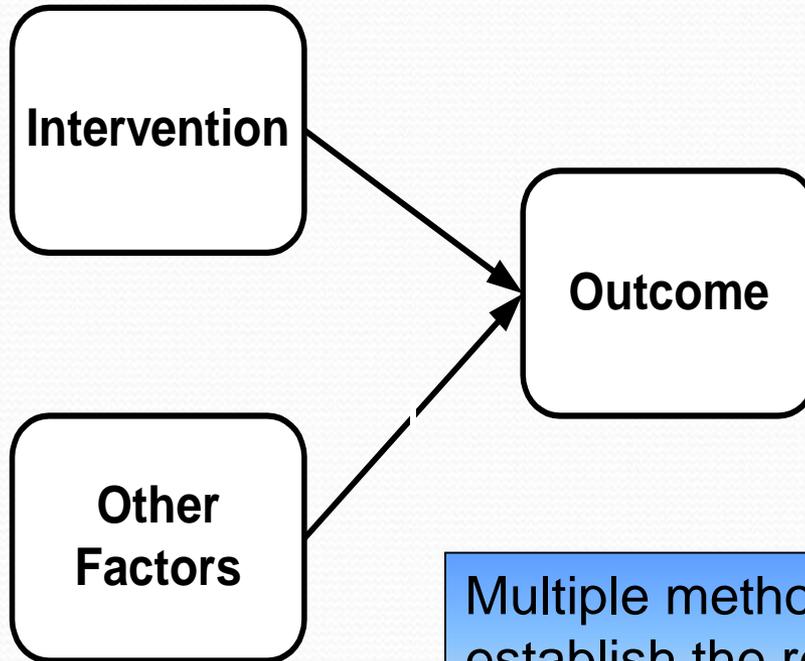
$e_1$  and  $e_2$  as above

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

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

$$X_1 = c_1X_2$$

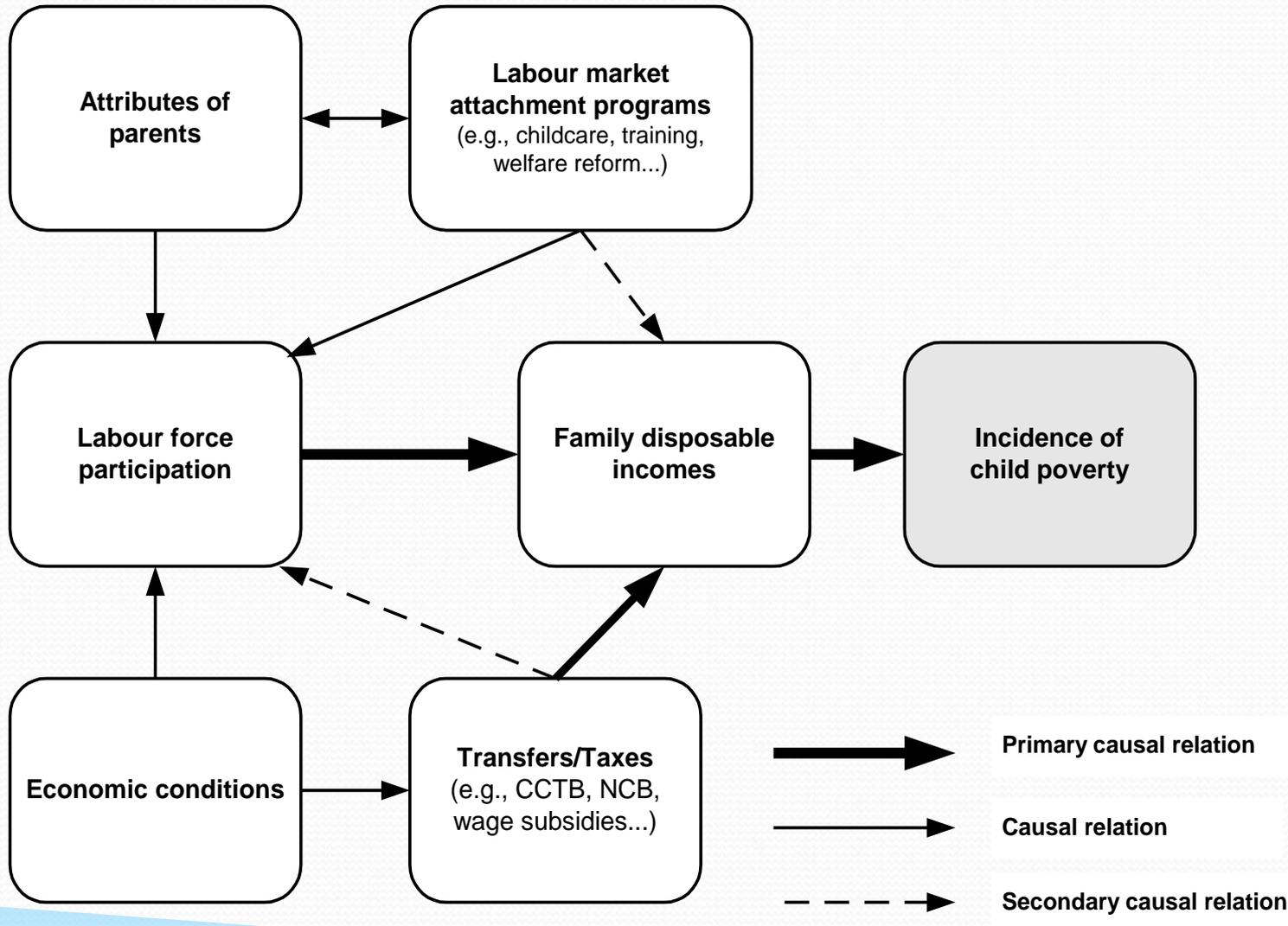
# Returning to causal logic models



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

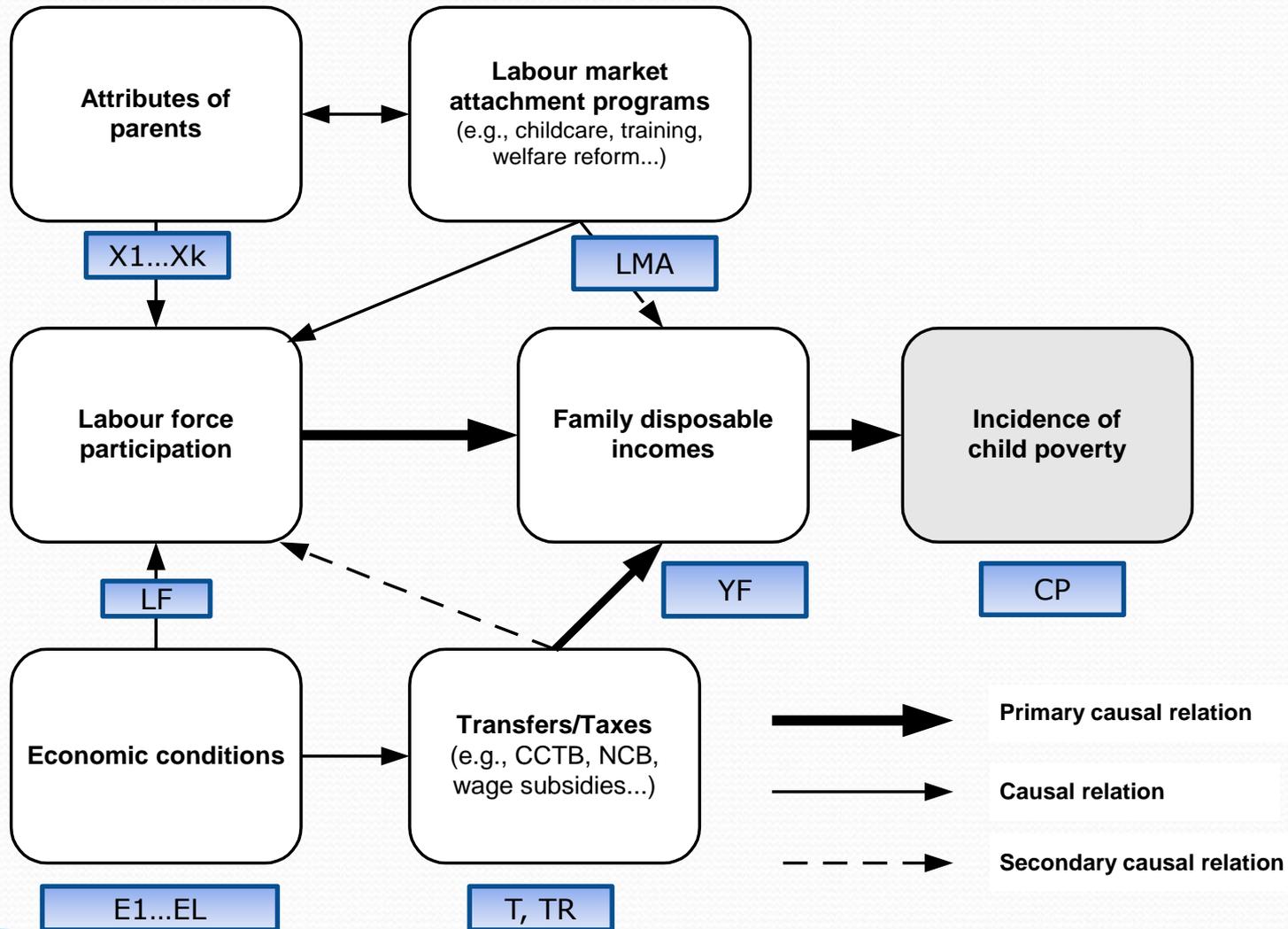
Multiple methods and experimental techniques establish the relative importance of causes of changes in outcomes

# Graphical logic for the National Child Benefit



# National Child Benefit

## Abstract theory of change logic model



## The “mechanism” of the NCB

The theory of change logic model supports the creation of an abstract or symbolic model

$$C_P = f(Y_F) \quad \dots 1$$

$$Y_F = f(LMA, T, TR, LF) \quad \dots 2$$

$$LF = f(X_1 \dots X_k, E_1 \dots E_L, LMA, T, TR) \quad \dots 3$$

$$(X_1 \dots X_k) = f(LMA) \quad \dots 4a$$

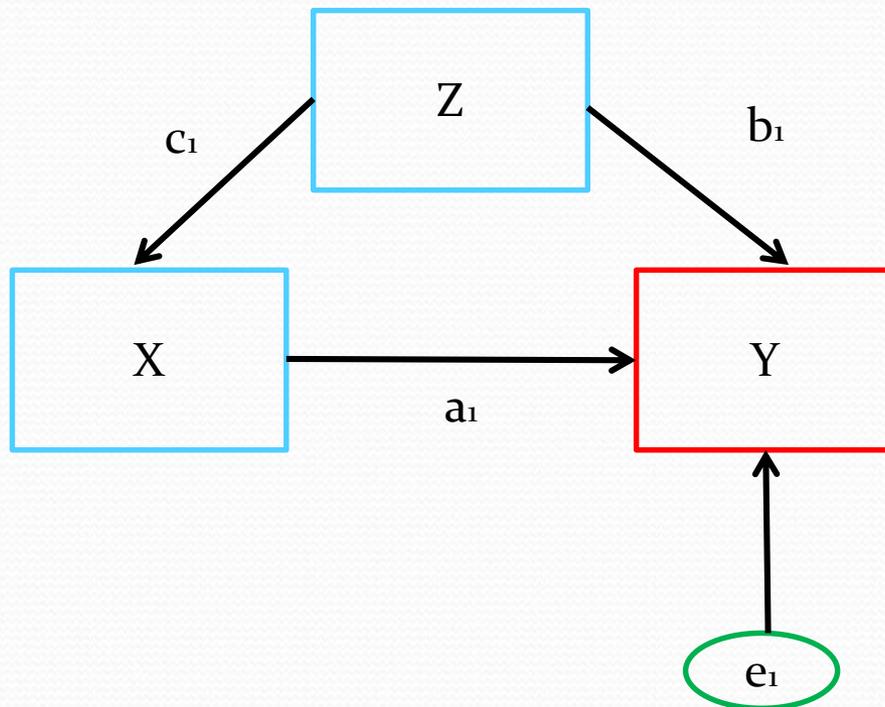
$$LMA = f(X_1 \dots X_k) \quad \dots 4b$$

The last equations (4a and 4b) are problematic since they suggest mutual determinism. The entire system requires more advanced modelling than simple regression.

$$TR = f(NCB, \text{Other (EI, CPP disability, In-kind, **adj. SA**)}) \quad \dots 5$$

## Causal analysis III: Confounding

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



$$Y = a_1X + b_1Z + e$$

$$X = c_1Z$$

Implies that

$$Y = a_1X + c_1Z + e$$

The problem is that the relationship of interest is  $X \rightarrow Y$ ; the confounding variable Z gets in the way in separating:

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

## Part 2 – Applied attribution analysis

## Causal Framework for Policy Design

### Thought experiments

- Theory of change reflected verbally or mathematically
- Scenario design and simulation
- Response curve analysis shows the hypothetical results based on a series of systematic trials
- Very helpful in developing a program.
- If the parameters of the model do not have an empirical basis, the model results can be very misleading.

### Observational studies

#### Statistical control and natural experiments

- The counterfactual exists as a dummy variable in the context of a standard multivariate model
- Sometimes separate regressions are run on the treatment and control groups.
- Statistical significance and magnitude on the counterfactual variable measures the net impact.
- Data sources include administrative files, client surveys and large datasets (SLID, SA, etc.).
- Natural experiments exploit unique one-time opportunities.

#### Propensity matching

- Creates a synthetic program and comparison group often from participants and non-participants.
- Various tests are used to compare the similarities of program and comparison groups.
- Assuming these tests are satisfied, the net impact is simply the difference between program and comparison group outcomes.
- Large datasets (SLID, SA, EI, ...) are often the source of information.
- Client data and surveys may be used to augment administrative data.

### Empirical experiments

#### Lab

- Limited randomization of participants
- Non standard populations
- Repeated manipulation of program parameters
- Control on all aspects of the experiment to maximize internal validity
- Net impact are directly measured as variations in response experimental manipulations
- Financial incentives (actual payment)
- N= 10 with many replications
- Results in days

#### Field

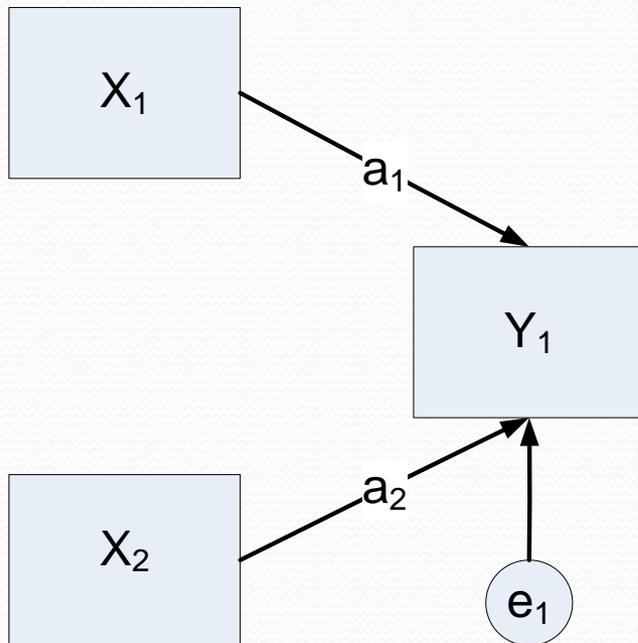
- Randomisation of participants (but this varies)
- Participants resemble the target population
- Less control over the experiment to increase external validity.
- Financial incentives usual but not mandatory
- N=200 - 500 with some replication
- Results in weeks/ months

#### Social experiments

- Replication of a clinical trial with large randomly selected sample
- Attempt to recreate "real world:" policy context.
- Policy (with full financial and administrative feature of the program.
- N=2500+
- Results in years

# Treatment effects

- “Treatment effect” is the effect of a given treatment or intervention on an outcome variable of interest.
- In the simple regression model



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

where  $X_1$  is the 0-1 policy variable, the treatment effect is the coefficient  $a_1$ .

# Models for estimating treatment effects

1. Randomized control (RC) (social experiments)
2. Regression models (regression discontinuity, difference in differences)
3. Matching estimators
4. Instrumental variables

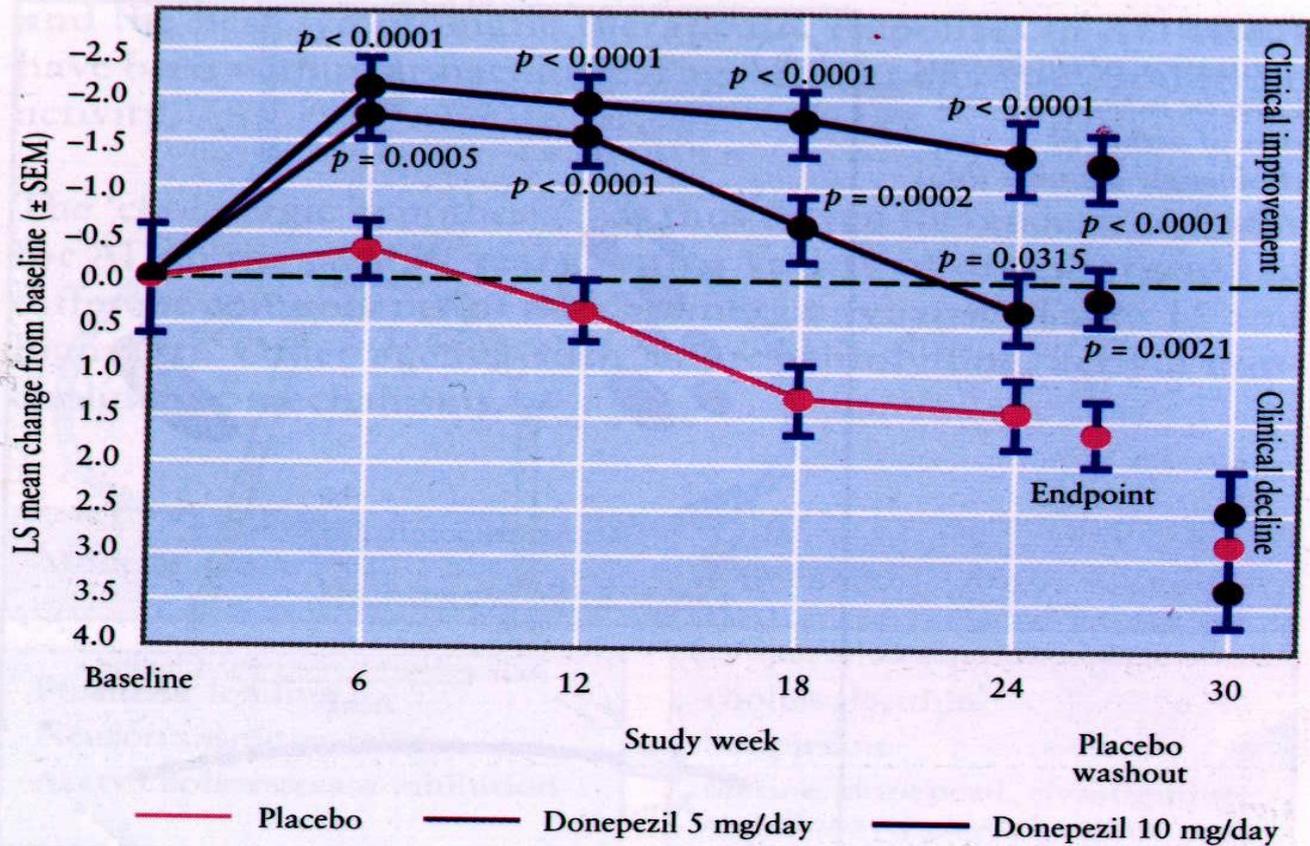
# Random experiments

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 subject receives.

# 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.
- Randomization includes a variety designs.



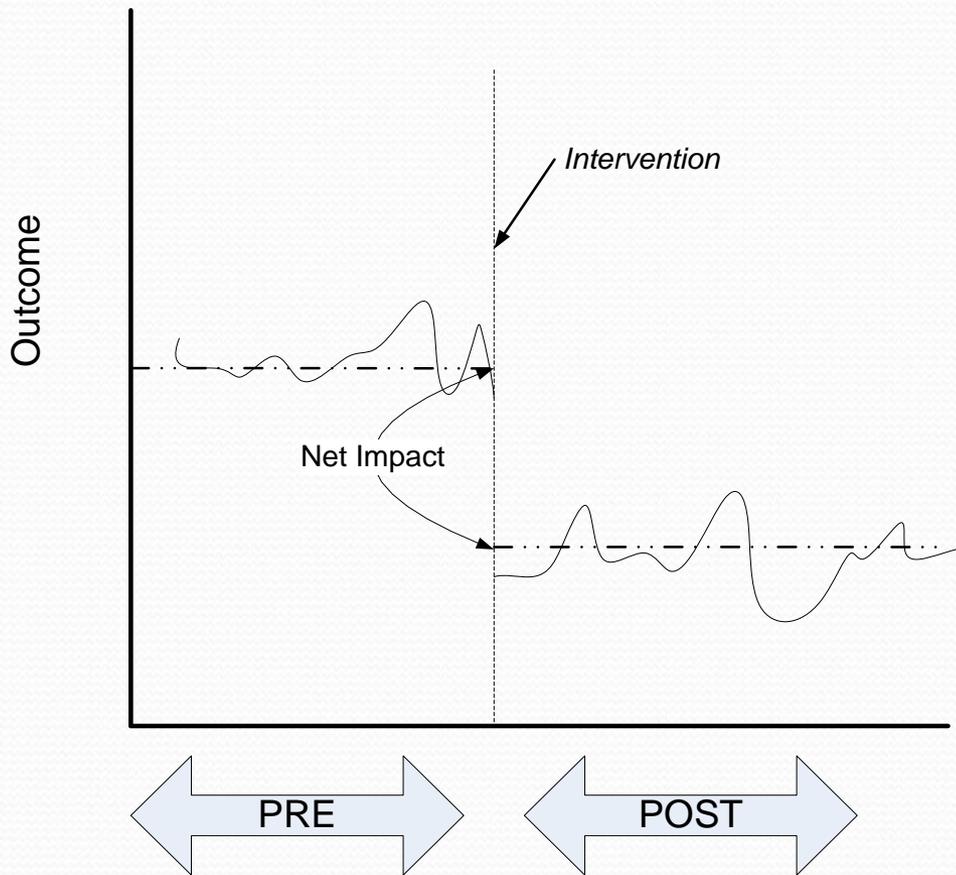
**Figure 10**  
*Illustration of therapeutic response to donepezil in an international double-blind, placebo-controlled trial, using the ADAS-cog at regular intervals followed by a 6 week placebo washout period (Gauthier et al 1998).*

# Limits of randomized designs

In social science, randomized double blind experiments are often not feasible for the following reasons:

- 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).

## Regression discontinuity — Basic pre-post design



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

Two common problems are:

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

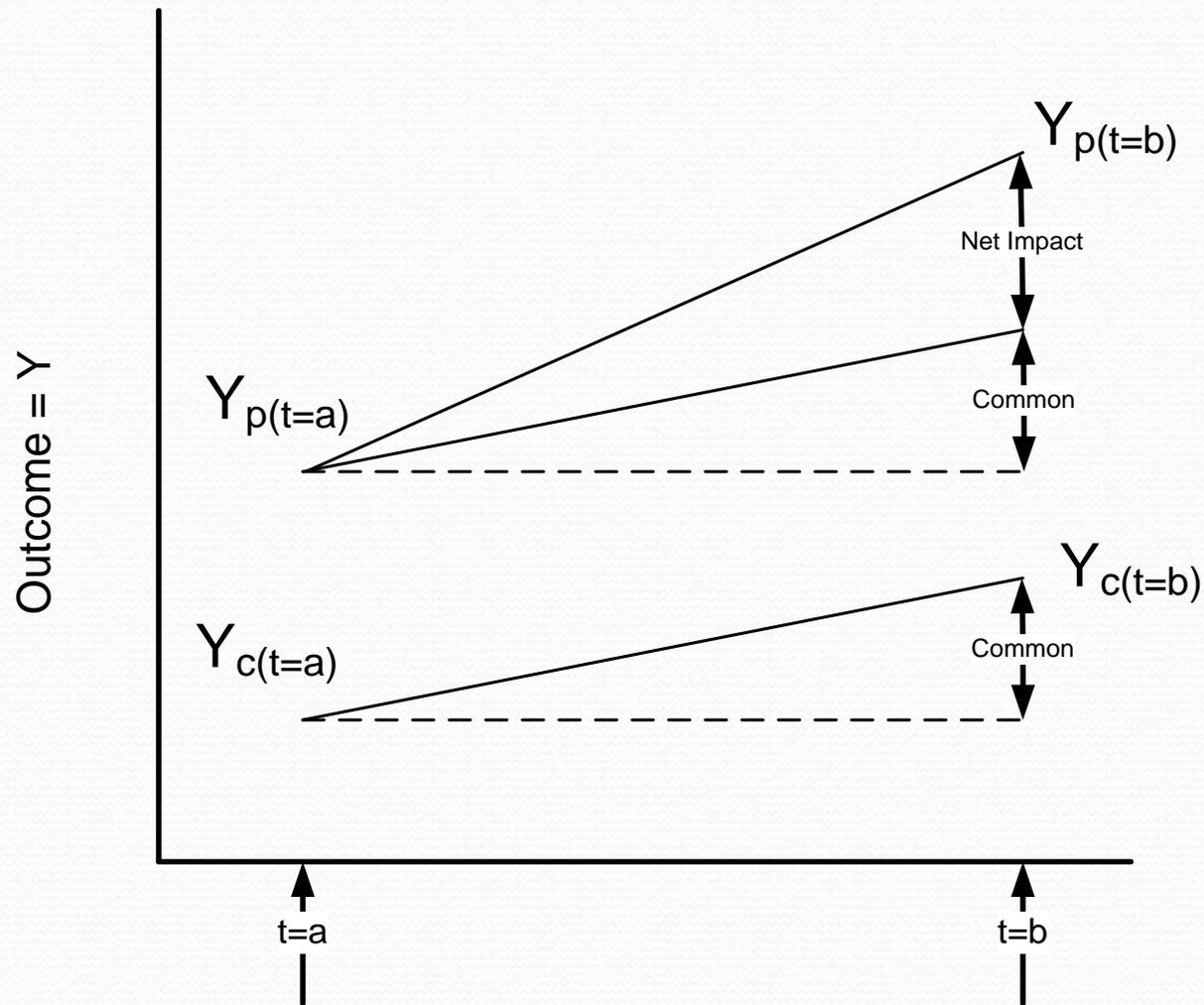
# Natural experiments

- Create a “split” in the sample, where treated and untreated are classified by a variable that is not related to the the treatment.
- This split occurs “naturally” where the program change occurs in one area/jurisdiction and 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.

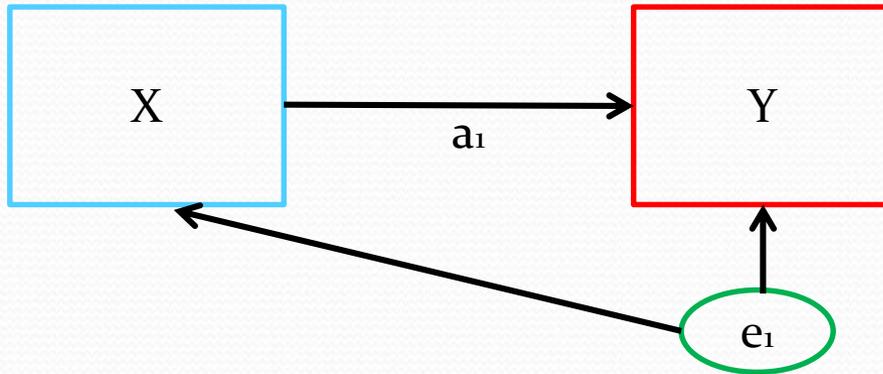
# Difference in Differences





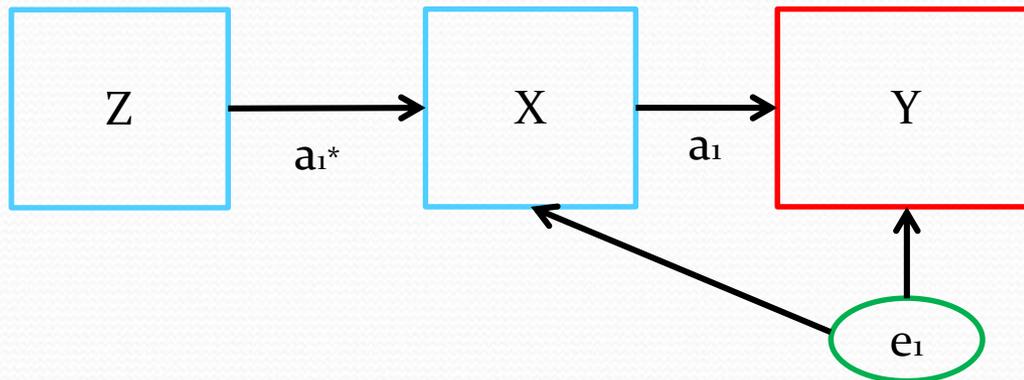
## A problem – When e confounds X: Instrumental variables

A core assumption of regression is that e does not affect X (X is measured without error)



$$Y = a_1X_1 + e$$

But what if  
 $X = f(e)$ ?



This model is termed  
“instrumental” variables.  
Z is the instrument.

# Matching

In social experiments, participants differ from non-participants because of the following:

- failure to hear of program
- constraints on participation or completion
- selection by staff

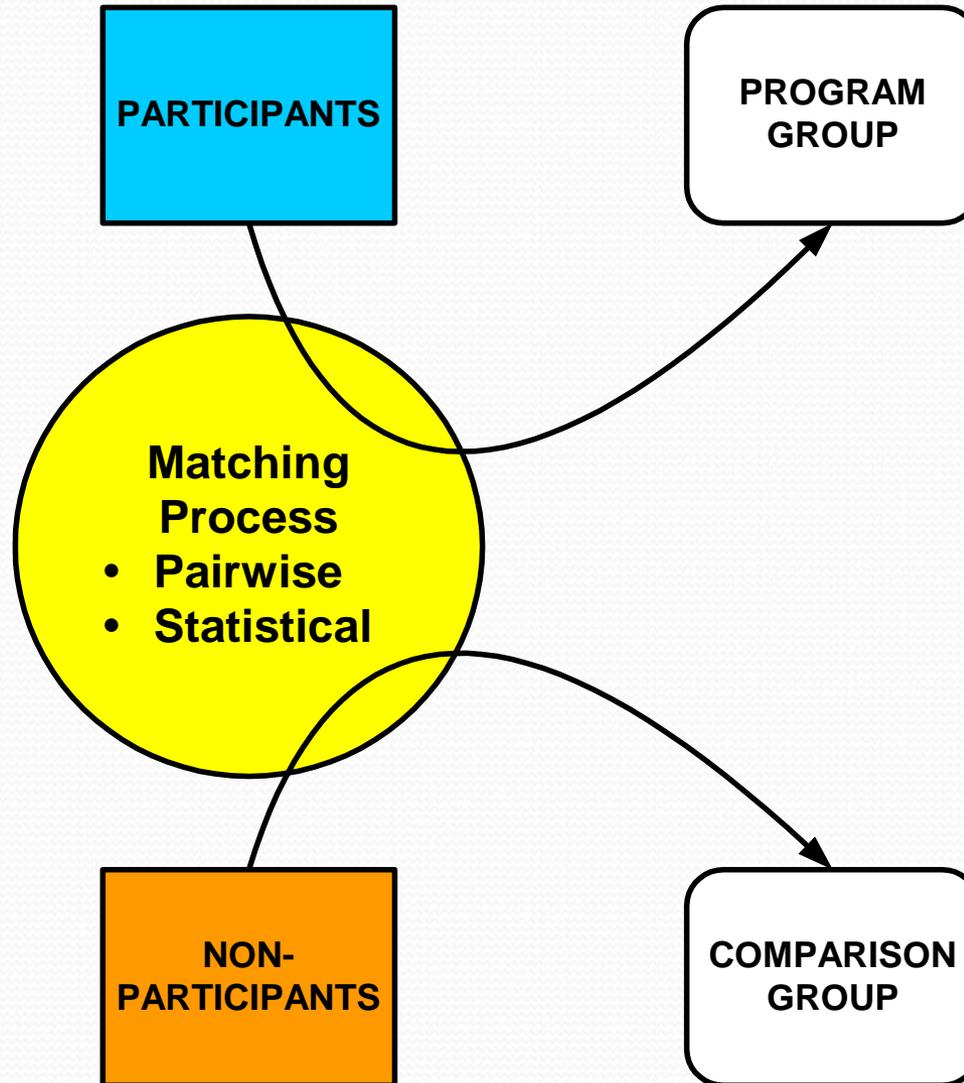
Creating a matched sample of participants and non-participants can be accomplished via the following:

- 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*.

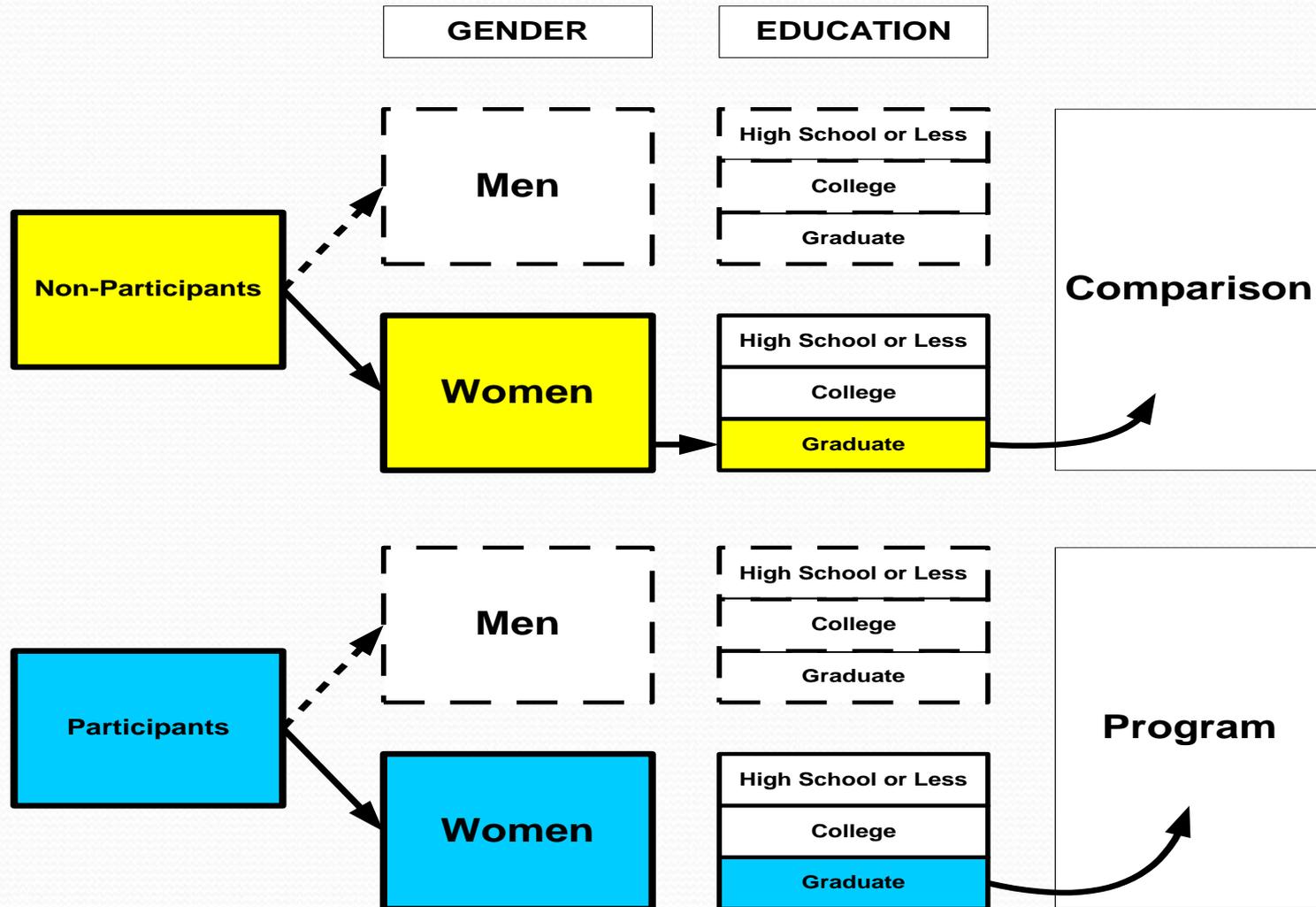
# Matching process



## 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.

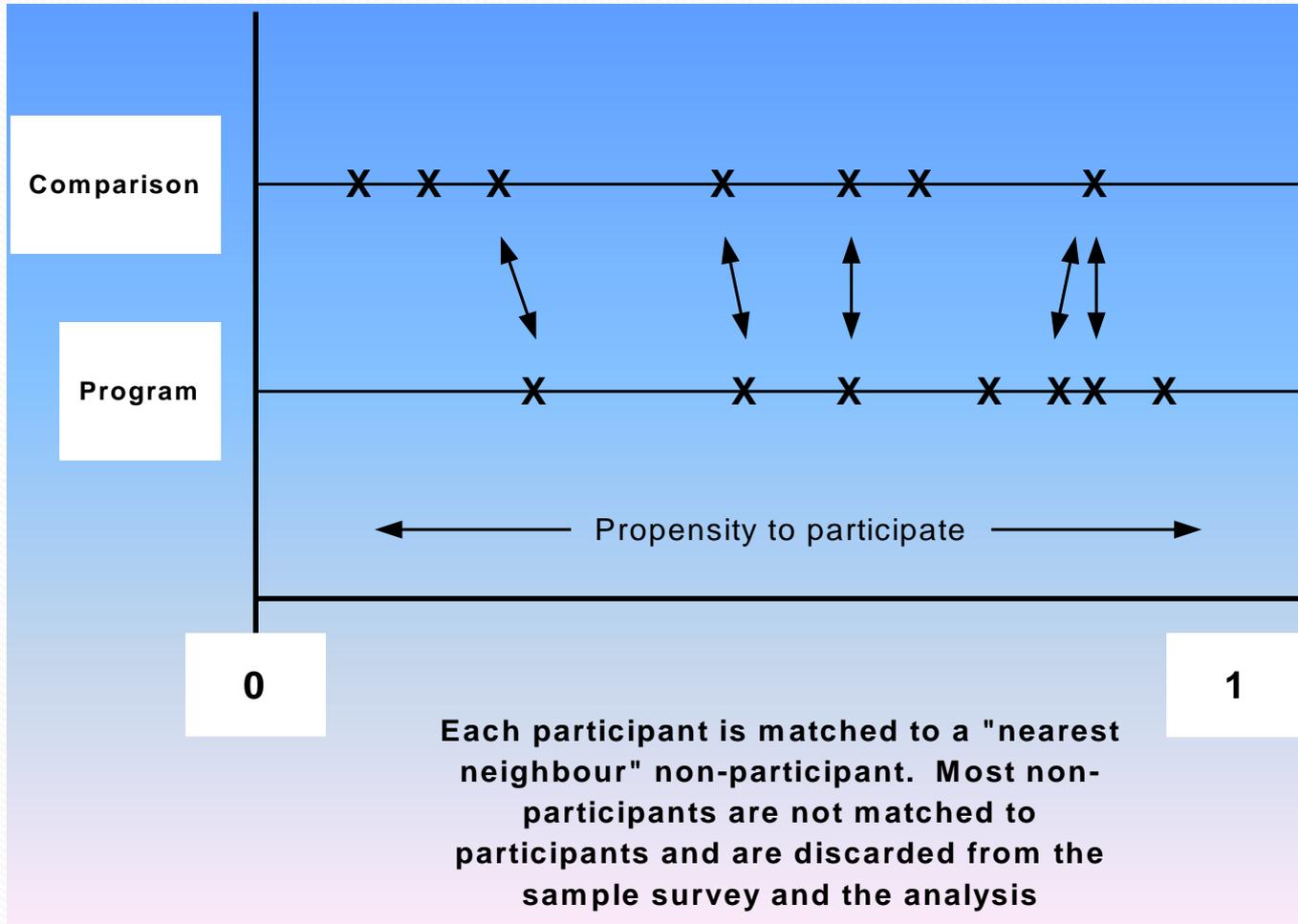
# Pair wise matching



# Issues in matching

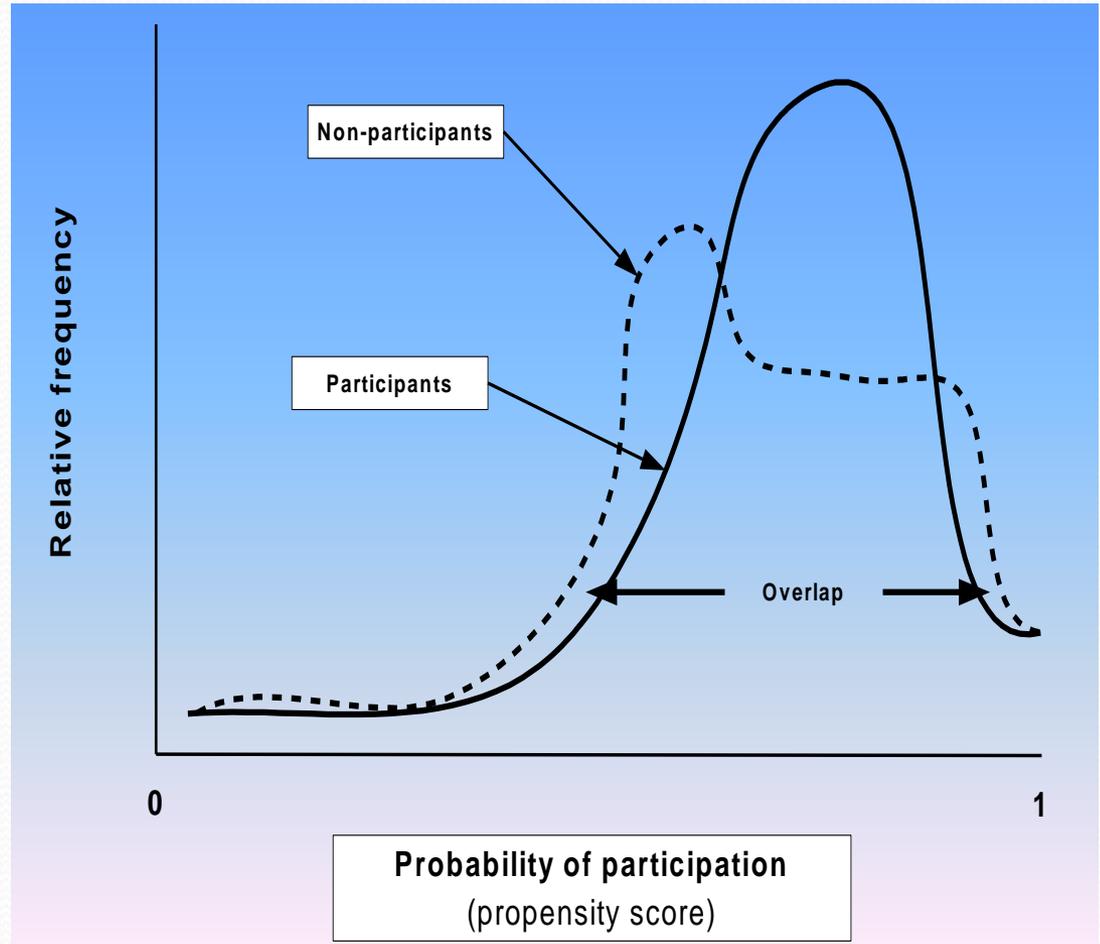
- The matching is limited to the variables available in the administrative files.
- Important missing variables such as age and number of children and incomes of other household members weaken the match.
- The matching produces samples that are statistically similar with respect to the matched variables.
- The key idea is that matching on the observed variables may not align the program and comparison groups on the non-observed variables.
- Every additional variable that is introduced to the matching process potentially improves the closeness of the match.

# Matching simplified

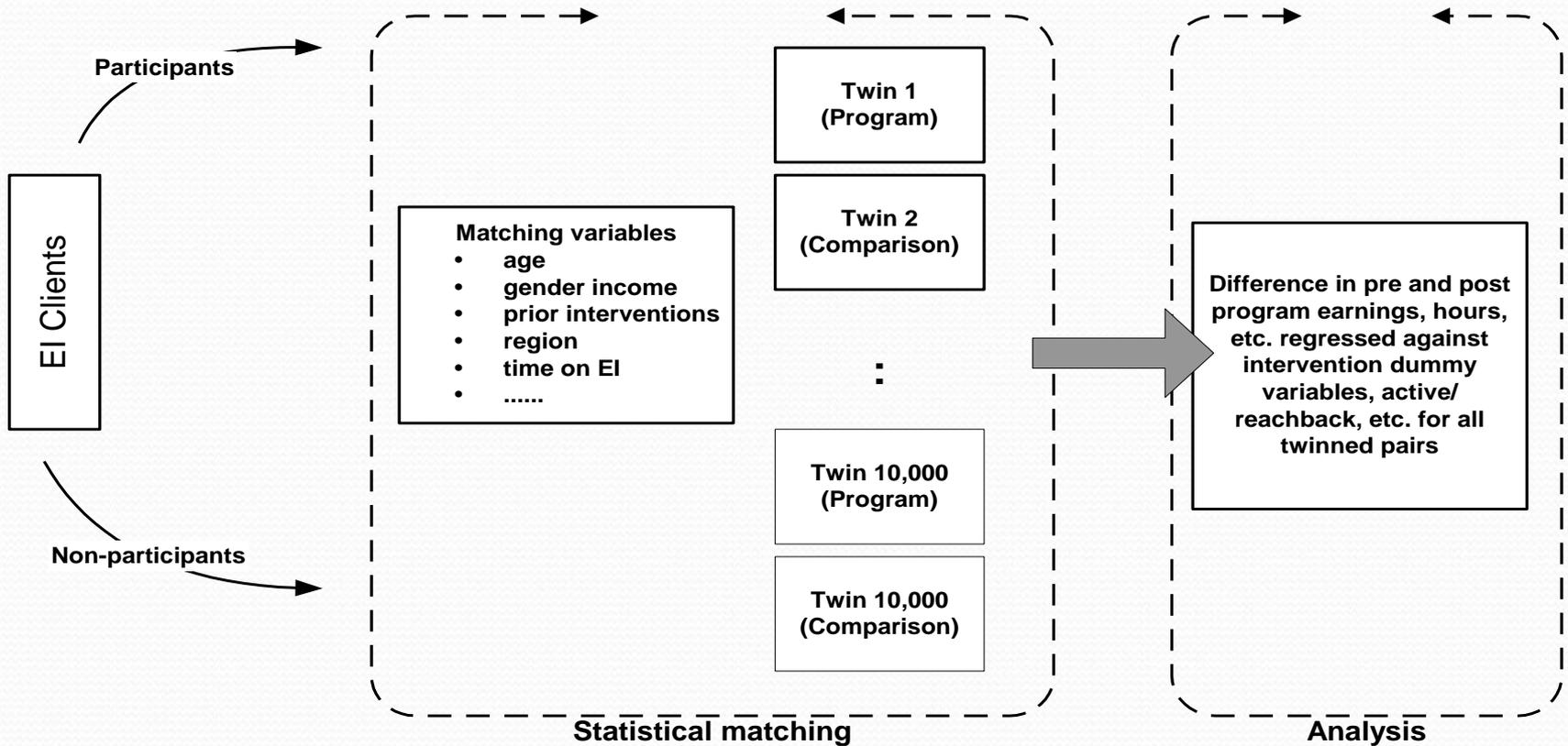


# Region of common support

- Each participant has the value of 1 for P, and each non-participant has the value 0.
- However, once the model is estimated, each participant and non-participant has a score between 0 and 1. Participants tend to have scores closer to 1 and non-participants are closer to 0.
- The distribution of scores can be graphed.



# Statistical matching applied to the LMMA



Statistical matching and structural modelling

## Part 3 - Examples

## Example 1 - RCA – Fluorouracil Incident

“On August 22nd, 2006, a 43 year old woman died after a medication incident that occurred while she was receiving outpatient care at the XXXX Cancer Institute in XXXX.

The cause of death as determined by the coroner was ‘sequelae of fluorouracil toxicity.’

On July 31, the woman had inadvertently received an infusion of fluorouracil over 4 hours that was intended to be administered over 4 days.”

ISMP Canada, *Fluorouracil Incident Root Cause Analysis*, <http://www.ismp-canada.org/download/reports/FluorouracilIncidentMay2007.pdf>

The Canadian Patient Safety Institute created a team of treatment specialists, human factors scientists, and other experts to deconstruct the treatment process, the accident, and the response.

## Example 2: Active labour market programs for disabled persons

- CM-LMAPD is a portfolio of programs intended increase the labour market participation of persons with disabilities
- It closely resembles the employment benefits and support measures developed by HRDC (HRSDC) and provinces/territories over the last 15–20 years
- One line of evidence in the evaluability assessment was to explore the feasibility of measuring the impact of training on a subset of the “disability sample”
- We used SA and training records (merged) to measure the impact of training on employment outcomes.