HM Treasury Seminar
Raising Standards in Policy Evaluation
7 October 2008

RCTs and Other Ways of Establishing a Counterfactual

Philip Davies, PhD
American Institutes for Research
Oxford and Washington DC

AMERICAN INSTITUTES FOR RESEARCH®

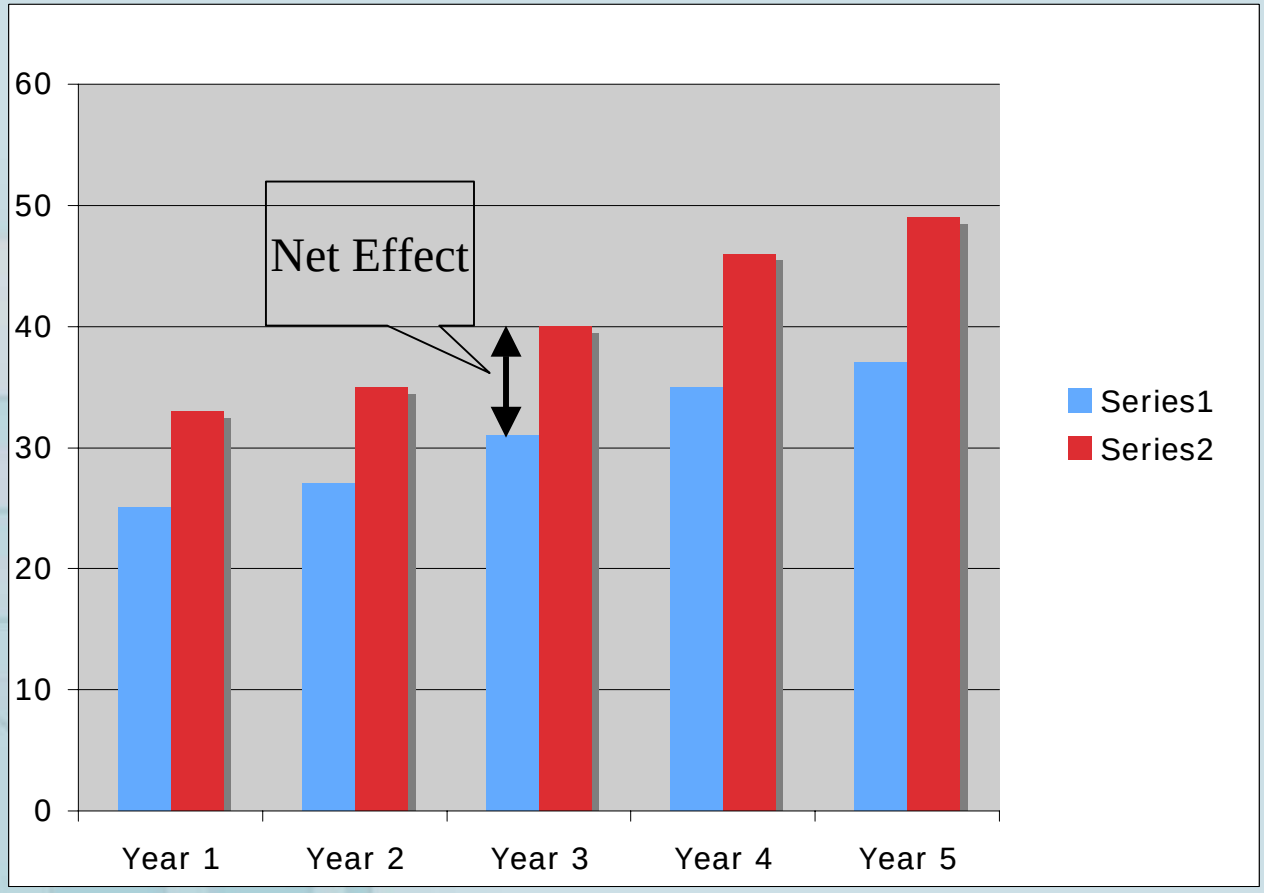
Effectiveness of What?

- *Intervention* effectiveness - what works?
- *Resource* effectiveness - at what cost/benefit?
- *Diversity* of effectiveness across different groups – what works for whom and when?
- *Implementation* effectiveness - how it works?
- *Experiential* effectiveness - users' views

The Importance of a Counterfactual (Establishing Net Effects)

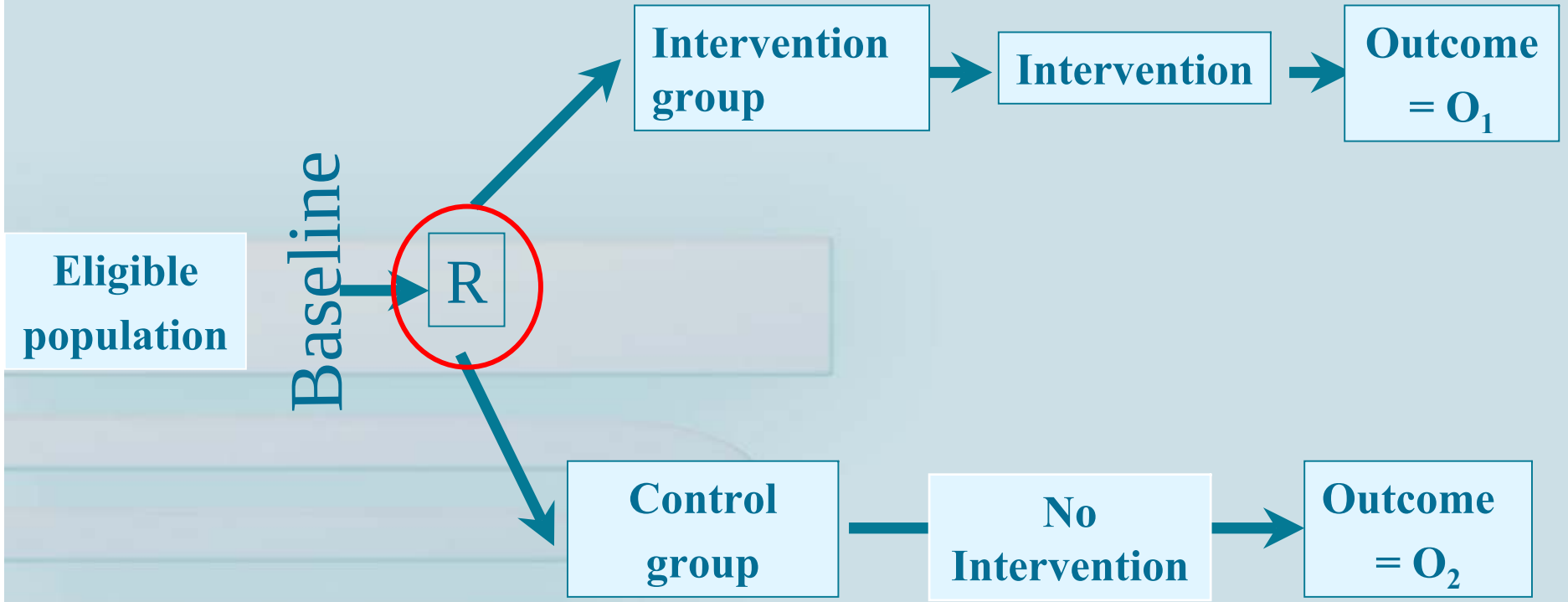
- Counterfactual: what would happen without the intervention i.e. doing something else or nothing at all
- Need to *estimate* the counterfactual
  - i.e. find a control or comparison group
- Ideal Counterfactual Criteria:
  - The intervention and the counterfactual groups should have identical characteristics on average,
  - Only reason for the difference in outcomes is due to the intervention

Evaluations of Net Effects (Against a Counterfactual)



American Institutes for Research®

Randomised Controlled Trial/ Random Allocation Experiment



Effect estimate = 'O_1-O_2'-
Counterfactual is O_2

AMERICAN INSTITUTES FOR RESEARCH®

Randomised Controlled Trial/ Random Allocation Experiment

- Gives each eligible unit/individual an equal chance of receiving the intervention, i.e.:
- Matter of chance as to who receives the intervention
- Or, for who receives the intervention first
- The only difference between the experimental and control groups is exposure to the intervention
- Random allocation ensures no systematic differences between the intervention and control groups in factors, known and unknown, that may affect outcome.

Randomised Controlled Trial/ Random Allocation Experiment

Requires:

- Allocation to be *independent* of service or policy providers
- Sufficient sample size and statistical power to identify a minimal detectable effect
- 'Blinding' or 'Double blinding' ➔ not always possible in public policy/public service delivery
- Meticulous project management

Impact Evaluations - Randomised Controlled Trials in UK

Department for Work and Pensions (DWP):

- Restart Programme (1990)
- Benefits Agency Visiting Officer pilot
- New Deal programme for people aged 25+
- New Deal for Lone Parents In-Work Training Grants
- Employment Retention and Advancement Demonstration (ERA)
- Job Rehabliation and Retraining Pilot (JRRP)

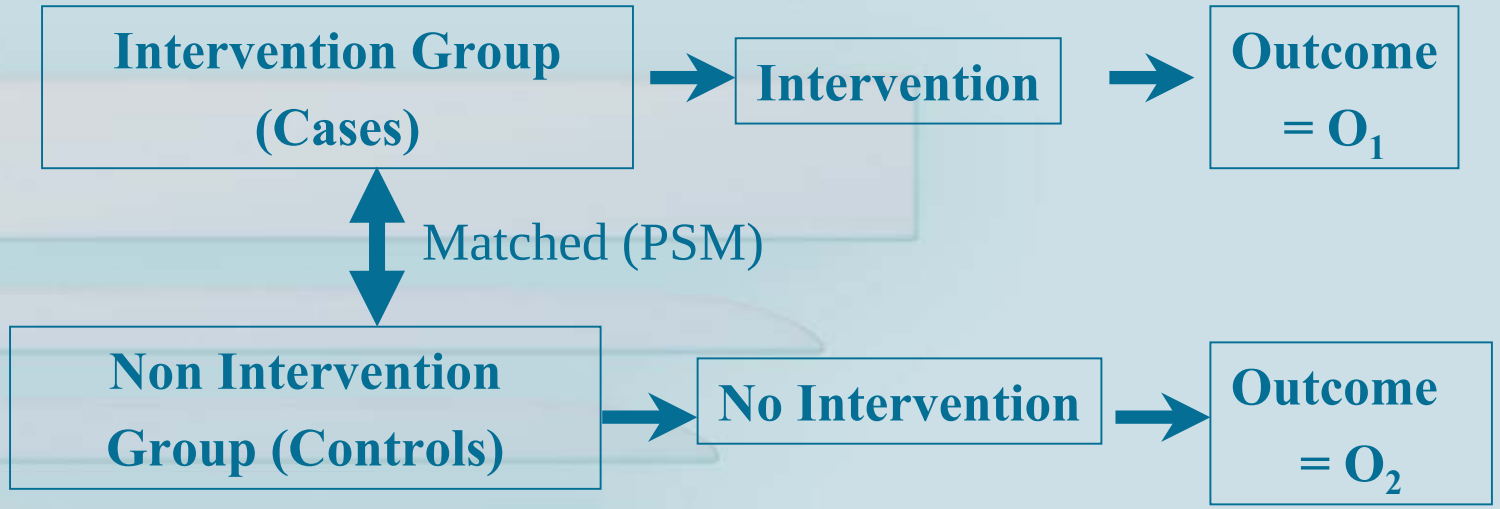
Home Office:

- Restorative Justice Project

Department for Education and Skills (DfES):

- Peer-Led Sex Education in Secondary Schools

Matched Comparison Design
Two Group Before and After-Studies/ Case Control Studies



Effect Size = O_1 - O_2

Note: Counterfactual is O2

Problems with Matching

- Cannot be participants against non-participants
- There may be some systematic differences between Ps and Non-Ps (e.g. ability, motivation) that explains participation/non-participation
- On what variables do you match?
- e.g. School Attainment: age? gender? family income? parents' educational attainment? housing status? nutrition and diet? etc.

Solution:

- Propensity Score Matching

Propensity Score Matching

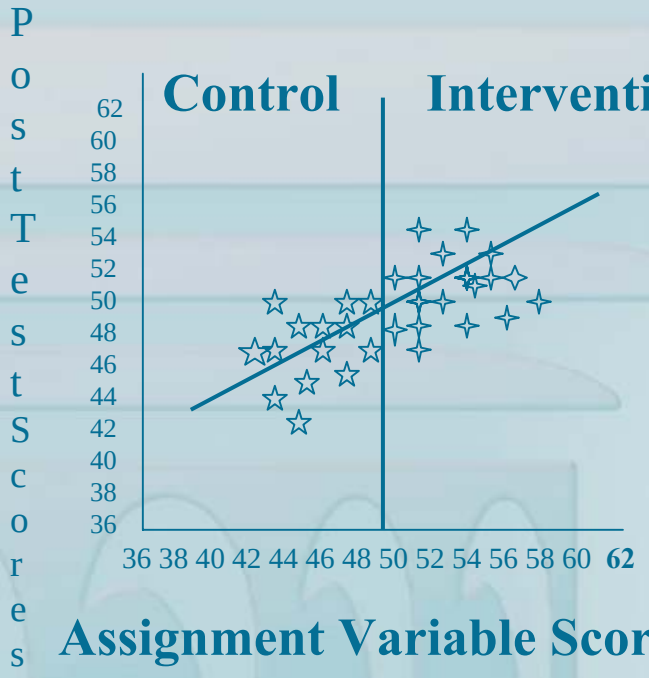
- Match on an *aggregate* score of individuals' likelihood (or propensity) to achieve at school
- Can be exact propensity score matching
- Or nearest neighbour
- Or kernal matching

Impact Evaluations - Matched Comparisons Designs

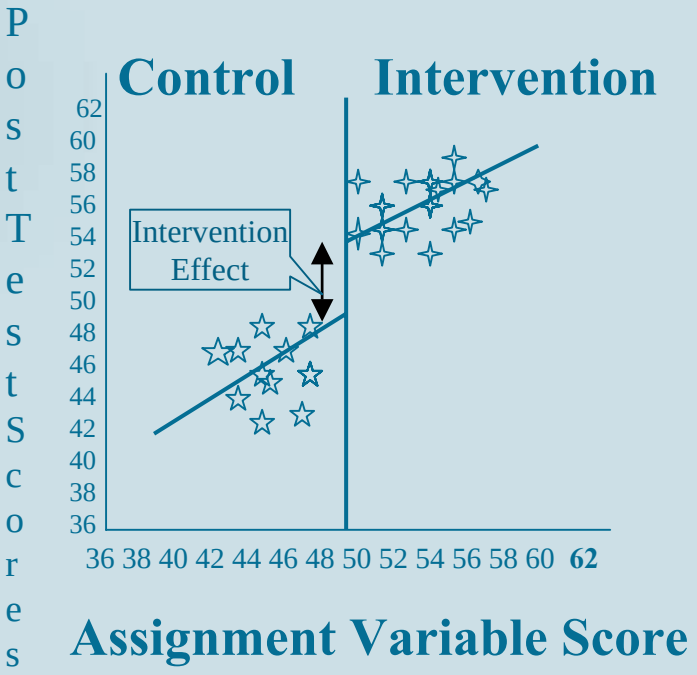
- Also used in UK government policy evaluations:
- DfES evaluation of Educational Maintenance Allowance (PSM)
- DWP evaluation of Work-Based Learning for Adults (PSM)
- Home Office evaluation of Cognitive Therapy for Offenders (Not PSM)
- DWP evaluation of Employment Zones (Not PSM)

Regression Discontinuity Design

Regression Discontinuity Trial With No Treatment Effects



Regression Discontinuity Trial With an Effective Treatment



Regression Discontinuity Design

- Assignment to the intervention is based on a clearly defined index, scale or parameter with a known cutoff for eligibility
- Can often take advantage of an official criterion for assigning people to the intervention (e.g. a benefit rule)
- No need to "exclude" a group of eligible households/individuals from the intervention for research purposes
- RD yields an unbiased estimate of treatment effect at the discontinuity
- However, the estimated impact around the cutoff may not generalize to entire population

Regression Discontinuity Design

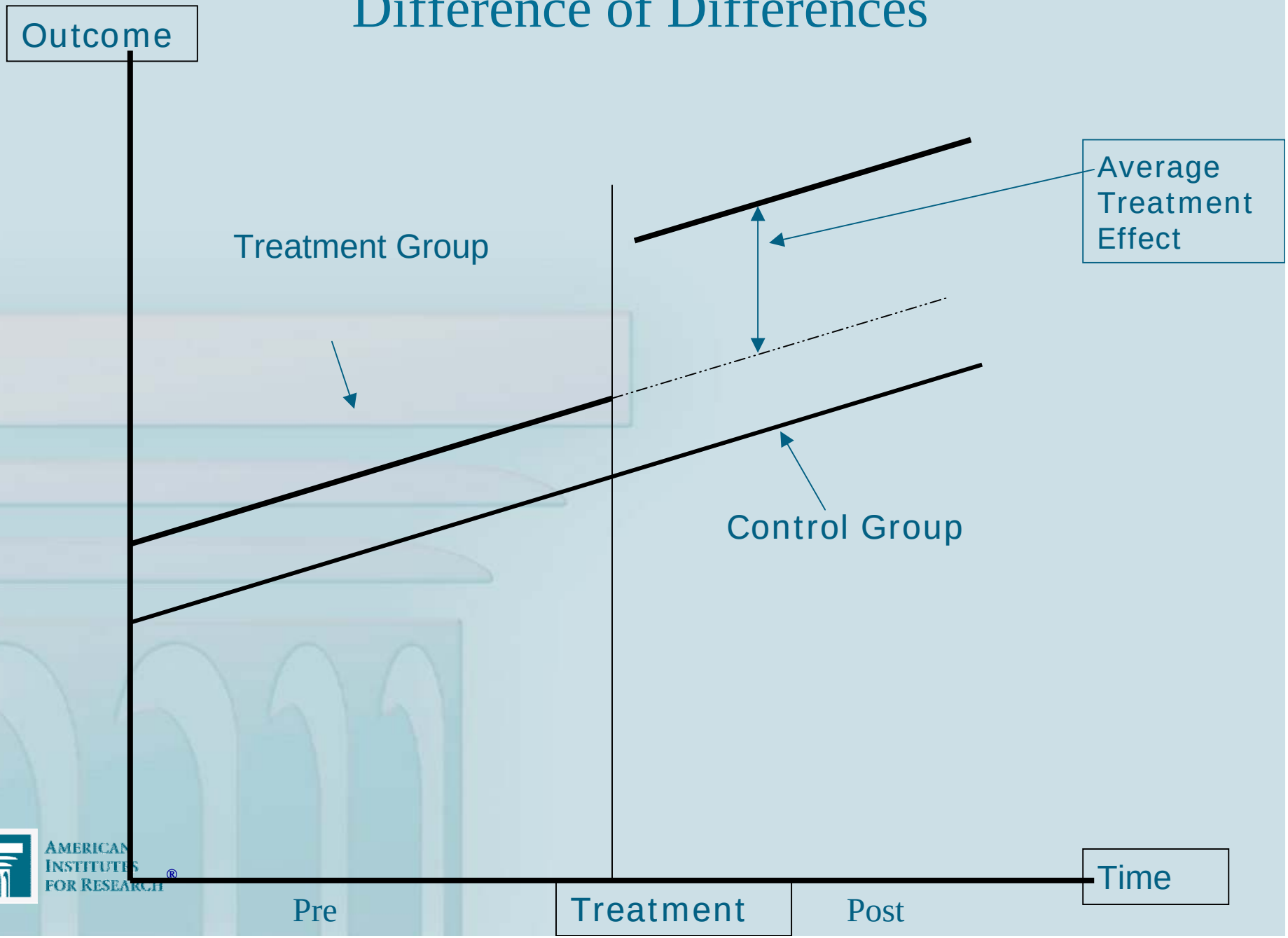
Examples with Targeted Social Programs

- Anti-poverty programmes are targeted to households below a given poverty index
- Pension programmes are targeted to population above a certain age
- Scholarships are targeted to students with high scores on standardised test
- Drug Treatments are allocated to offenders according to their scores on a re-offending likelihood scale

Difference of Differences

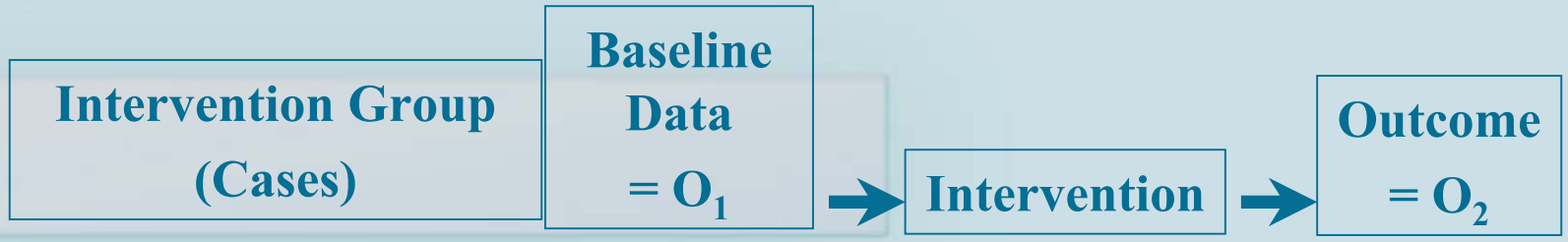
- Compares data on treatment and control group *before* the treatment to estimate the *'normal' difference* between treatment and control group
- Then compare this with the *actual difference after* the receipt of treatment.
- The difference between the actual post-treatment data and the estimate of the 'normal' difference (the trend line) is the average treatment effect

Difference of Differences



AMERICAN INSTITUTES FOR RESEARCH®

Single Group Before and After Studies/ Cohort Studies



Effect Size = $O_2 - O_1$

Note: There is no counterfactual

- Difference may be due to extraneous variables, or regression to the mean

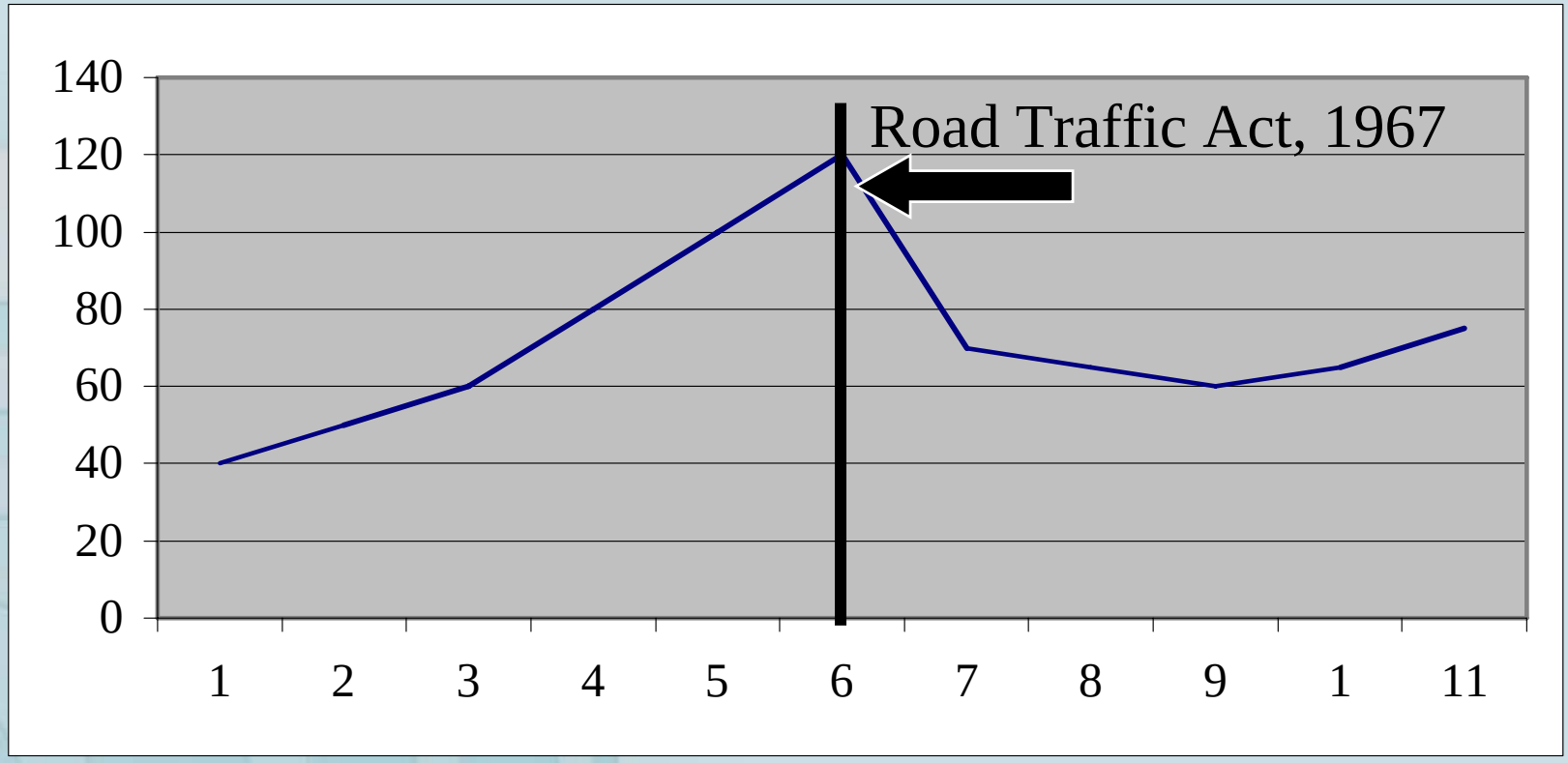
Warning: DO NOT USE

Interrupted Time Series Design

- Data consist of a time series in the form of a sequence of baseline or routine measurements
- Interruption to the time series may be attributable to an independent causal variable
- All possible causal factors/rival hypotheses must be rejected until the candidate independent variable can be invoked.
- Some degree of speculation and data fishing can be involved with ITS designs

Interrupted Time Series Design

Road Traffic Deaths, UK, 1950-1972



Interrupted Time Series Designs

- Evaluation of the impact of the Road Traffic Act 1967
- Many evaluations of medical and public health interventions
- Evaluation of literacy amongst primary school children
- Evaluation of alcohol licensing and crime
- Evaluation of street lighting and crime
- Evaluation of CCTV and crime

Lest We Forget

- RCT, RDD, PSM, ITS are simply ways of establishing a counterfactual
- We still need data:
  - Survey data
  - Administrative data
  - Process data
  - Qualitative data
  - Costs data
  - Outcomes data (positive and negative)
- Targets ≠Outcomes

Contact

Philip Davies PhD

Executive Director, AIR UK

2 Hill House
Southside
Steeple Aston
Oxfordshire OX25 4SD
Tel: +44 1869 347284
Mobile: +44 7927 186074
PDavies@air.org

1000 Thomas Jefferson Street, NW
Washington DC 20007
Tel: 202 403-5785
Mobile: 202 445-3640
PDavies@air.org

AMERICAN INSTITUTES FOR RESEARCH®