The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Evaluation designs for interventions aimed at reducing health inequities

Dr Puja Myles

Puja.myles@nottingham.ac.uk

Session learning outcomes

You should be able to:

1. Recognise common evaluation designs used for assessing the effectiveness of interventions
2. Apply evaluation designs for assessing the effectiveness of interventions specifically addressing health inequities

THE CONCEPT OF THE COUNTERFACTUAL

Counterfactual theories of causation

- Causal claims can be explained in terms of counterfactual conditionals of the form:

  "If A had not occurred, C would not have happened."

  (Stanford Encyclopedia of Philosophy)

- In the context of evaluation:

  If the intervention had not occurred, the observed outcome/change in outcome would not have occurred; therefore the observed outcome can be attributed to the intervention

Counterfactual quirks

- We cannot actually observe the counterfactual (the what if this hadn't happened) scenario[1] (unless you are in a fantasy Star Trek world with parallel universes)
- We can approximate what the counterfactual scenario would be like using various evaluation designs
- The counterfactual can be approximated using qualitative approaches and in-depth case-studies of intervention participants

[1] unless you are in a fantasy Star Trek world with parallel universes

BROAD TYPOLOGY OF EVALUATION DESIGNS

Experimental designs (ED)

- Compare health outcomes of interest in intervention group and non-intervention group (could be 'no intervention' or 'other intervention')
- Ideally, randomly allocate people to intervention and non-intervention group
- Random allocation aims to achieve 'balanced/ equivalent' groups that are similar in all respects other than receiving the intervention

Quasi-experimental designs (QED)

- Try to emulate the experimental design but due to feasibility or ethical constraints, cannot randomly allocate people to intervention/non-intervention groups (non-equivalent group design)
- Sometimes there may be no comparison group[1]
- Many variants of both experimental and quasi-experimental designs

[1] In these instances, measure change in outcomes over time (pre-intervention and post-intervention)

Are experimental designs methodologically better than QEDs?

- Randomised controlled Trial (RCT) considered the gold standard
- Good for demonstrating efficacy
- Poor predictor of effectiveness (translation to real world contexts)
- Not always appropriate for complex public health interventions
- Blinding not always possible with PH/HP interventions; contamination because of 'diffusion' can occur ; cluster RCTs possible

Internal validity

- Validity: the extent to which an evaluation actually measures what it sets out to measure

- Internal validity is essentially concerned with detecting change and being able to attribute it to the intervention

- Green and South (2006)

Common quasi-experimental designs

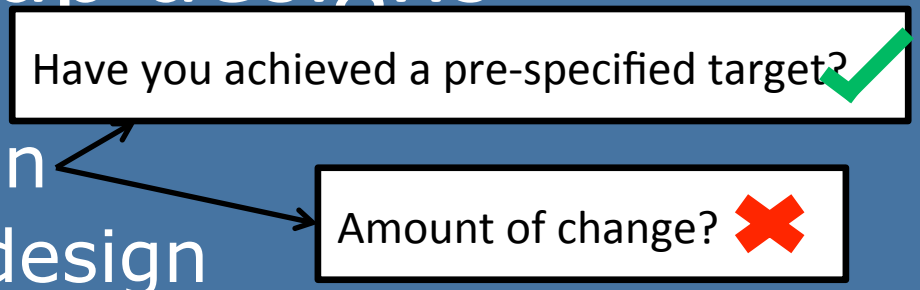
- Designs without a control group
- Designs with a control group but no pre-test
- Designs with both control groups and pre-tests

(Shadish et al. 2002)

Designs without a control group

- Also known as single group designs
- Strengths: Simple, inexpensive
- Subject to logical fallacy ‘post hoc ergo propter hoc’: after this, therefore because of this

Single group designs



- Post-test only design
- Pre- and post-test design

- Threats to internal validity?
  - History threats- historical or concurrent events
  - Maturation threats- participant maturation
  - Testing threats- pre-test acts as a stimulus for behaviour change
  - Instrumentation threats- familiarity with test instrument or differences in measurements

Designs with a control group but no pre-test measures

- Comparison group post-test only
- Can have one or more control (or comparison) groups
- Assumption: control group has similar baseline health status to intervention group
- Threats to internal validity: Selection bias (comparison groups are not comparable)

Designs with both control groups and pre-test measures

- Two group pre-test post-test
- Can have more than one comparison group
- Threats to internal validity?
  - Selection bias
  - Regression to the mean (RTM)
  - Multiple pre-test measures could help to deal with RTM (use average of pre-tests as your baseline)

Application of these basic designs to the evaluation of interventions aimed at reducing health inequalities

Key reference: Mackenbach, J.P. and Gunning-Schepers, L.J. (1996).
How should interventions to reduce inequalities in health be evaluated?
JECH. 51:359-364.

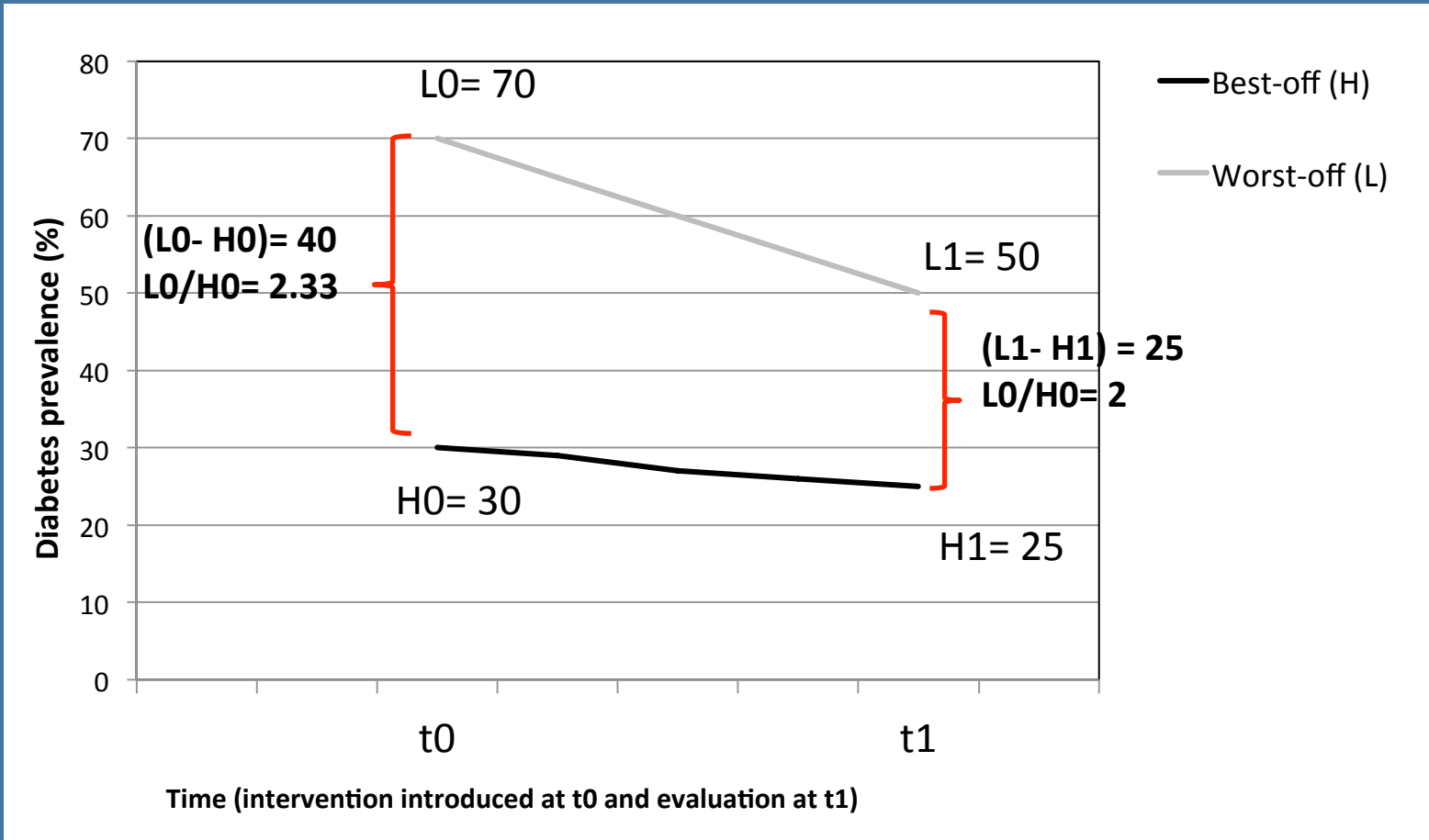
Assumptions

- Policy approach of narrowing the health divide between the best-off and the worst-off
- Reduction of SE inequalities by lowering rate of health problems in the lower SES group, not by increasing the rate of health problems in higher SES group

Case study

- **Problem:** Higher diabetes prevalence in low socioeconomic group as compared to high socioeconomic group
- **Intervention:** Free cholesterol monitoring and prescription of statins (cholesterol lowering agent) in low SES group (targeted intervention)

Hypothesis: Intervention effective if (L1-H1) < (L0-H0) i.e. reduction in absolute gap or if L1/H1< L0/H0 i.e. reduction in relative gap



Design C1: Single (intervention) group, pre-test and post-test measure, measures for both best-off and worst-off groups

REPRESENTING VARIOUS DESIGNS AND HYPOTHESES USING NOTATION

Design Notation[1]

		Before the intervention	After the intervention
Experimental condition (intervention)	Low socioeconomic status group	L0 (baseline health measure/status)	L1 (post-intervention health measure/status)
	High socioeconomic status group	H0	H1
X (random allocation)			
Control condition (no intervention)	Low socioeconomic status group	l0	l1
	High socioeconomic status group	h0	h1

[1] As per Mackenbach and Gunning-Schepers (1996)

Evaluation design options -1

Design	Hypothesis to be tested	Assumptions required
A: Two group, pre-test/ post-test; measures for both low and high SES groups	(L1-H1) – (L0-H0)< (l1-h1)- (l0-h0); [Post-intervention health inequality gap] – [pre-intervention health inequality gap] bigger in intervention group compared to control group	Experimental version considered gold standard; QED - account for confounding/bias
B1: Two group, pre-post; measures for low SES group only	(L1-L0) < (l1- l0); Reduction in adverse health outcome over time in intervention group is bigger than in the control group	(H0-H1)= (h0- h1); i.e. no change in health outcomes over time for the high SES group in both intervention and control arms
B2: Two group; post-intervention measures for low SES group only	L1< l1; Lower adverse health outcomes in intervention group compared to control group	As B1, plus L0= l0; i.e. baseline health status measures were similar in intervention and control groups

Evaluation design options -2

Design	Hypothesis to be tested	Assumptions required
C1: Single group; pre-post measures for both low and high SES group	(L1-H1) < (L0- H0); Post-intervention health inequality gap is smaller than pre-intervention health inequality gap in intervention group	(l1-h1)= (l0-h0); i.e. Health inequality gap remains unchanged over time for hypothetical control group
C2: Single group; pre-post measures for low SES group only	L1< L0; Lower prevalence of adverse health outcomes post-intervention in target group as compared to baseline prevalence in same group before implementation of intervention	As C1, plus: H1= H0; i.e., no change in health status of high SES group over time
D1: Two group; post-intervention measures for both low and high SES groups	(L1-H1)< (l1-h1); health inequality gap in intervention group is smaller than in control group	(L0- H0)= (l0-h0); i.e. at baseline, the health inequality gap was the same in intervention and control arms

Evaluation design options -notes

- Hypotheses postulated in terms of absolute differences in frequency of health problems between low and high SES groups (e.g. L1-H1); to frame hypotheses in terms of reduction in relative gaps, use ratio measures (e.g. L1/H1)

- Designs involving ≥2 groups could be either true experimental or quasi-experimental depending on random allocation

- These formulae assume that the health outcome measure relates to prevalence of a health problem; therefore, an effective intervention should decrease the prevalence of the health problem

OPTIONAL EXTRAS

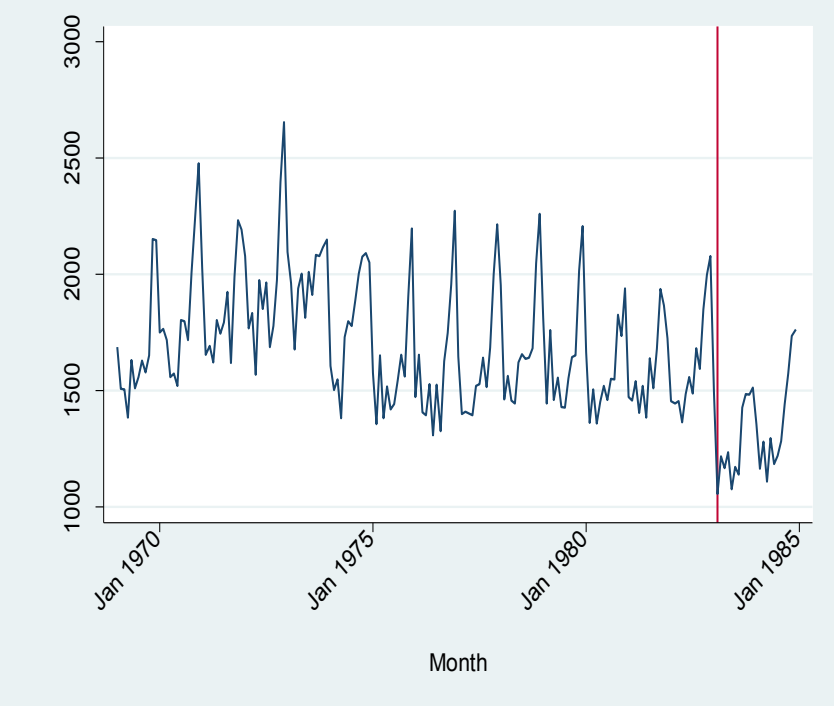
Using phased implementation to create comparison groups (stepped wedge designs)

- Useful for evaluation of new regional or national programmes when it would not be ethical to deny the intervention to any group
- Implement the programme in 2 or more phases (groups/ areas A, B, C).
- Baseline measures for A, B, C
- Phase 1: Implement in group A. B and C act as comparison groups.
- Phase 2: Extend programme to group B. Outcomes in A and B compared to control group C.

Time series design

- Used to quantify changes in an outcome measured over time, taking into account existing long term and seasonal trends.
- Can be applied to single or multiple comparison groups
- E.g. Did the introduction of a seat belt law in the UK affect the number of deaths and serious injuries in road accidents?

Did seatbelts have an effect?



- 31st Jan 1983 – compulsory for drivers and front-seat passengers to belt up
- Some long-term decline already?
- Strong seasonal pattern – more deaths & injuries on dark, short, winter days

Reproduced with permission from Dr Lisa Szatkowski, University of Nottingham

A useful design for interventions aimed at reducing health inequities

- Regression discontinuity design (RCD) allows you to target the intervention to those who are most needy
- Claimed to be the strongest non-equivalent group design and almost as robust as an RCT
- Sample size requirements greater than RCT as not as statistically powerful

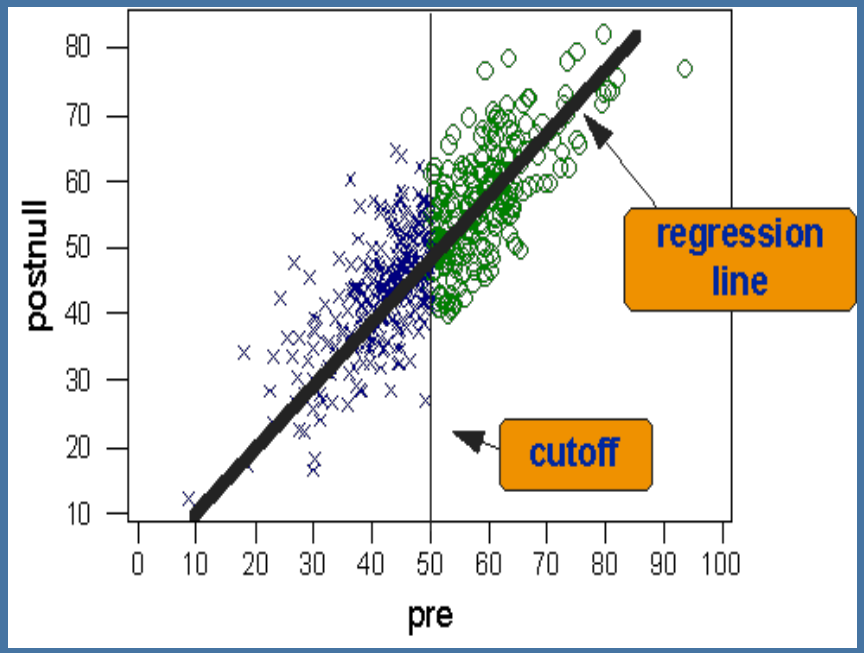
For a detailed discussion:

1. Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: <http://www.socialresearchmethods.net/kb/> (version current as of October 20, 2006).

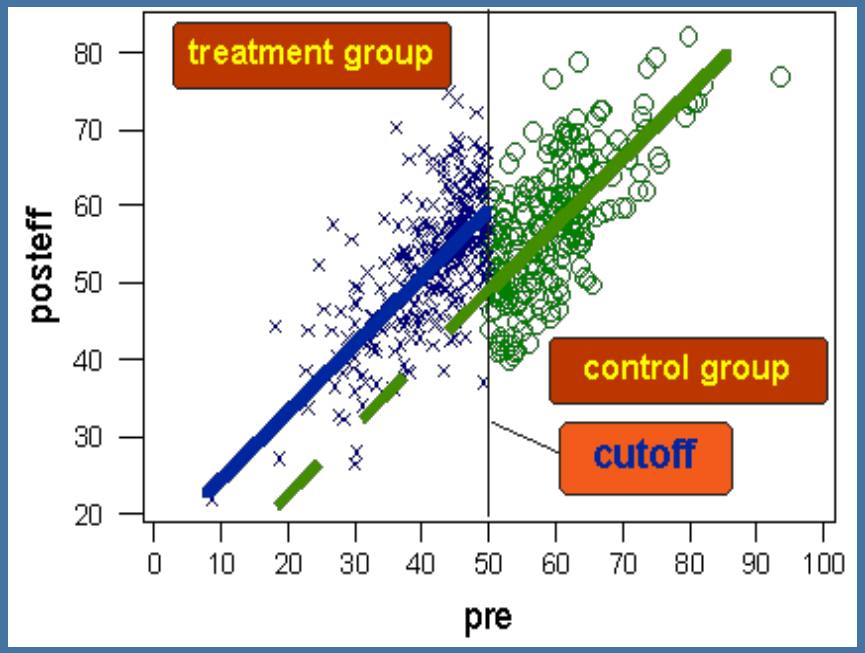
2. Chapter 7 in Khandker, S.R., Koolwal, G.B., Samad, H.A. (2010). Handbook on Impact Evaluation: Quantitative methods and practices . The World Bank.

The RD Design illustrated

In the absence of intervention



In the presence of intervention



Assumption in this illustration: desirable health outcome like life expectancy i.e. more is better

Reproduced with permission from Professor W. Trochim, author of the Research Methods Knowledge Base website (permission obtained 20.11.2014)

A note on generalised causal inference and external validity

- External validity is about generalisability
- Related to sampling: primary aim of sample selection is that we should be able to generalise our results based on that sample back to the population from which it was chosen

External validity vs. Internal validity in evaluation

"Greater weight should be placed on internal validity than on external validity because of what use is generalising a relationship if one doubts the relationship itself."

–Donald Campbell

"There are no enduring, context-free truth statements, and all human behaviour is time and context bound; we began to doubt seriously the possibility of generalisation from one site to the next."

-Guba and Lincoln