# Counterfactuals and Causal Policy Effects

## Advanced Social Epidemiology PhD Course

Sam Harper, Arijit Nandi

University of Copenhagen
2024-12-02 to 2024-12-06

# Causal Policy Effects

I. "What-If" Questions and Counterfactuals

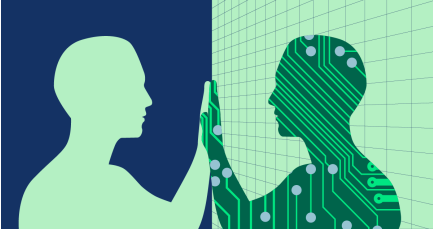II. Causal Parameters

III. Causal Assumptions

# Causal Policy Effects

## I. "What-If" Questions and Counterfactuals

## II. Causal Parameters

## III. Causal Assumptions

# "What-ifs" and counterfactuals

- Questions about the impact of an intervention (the change that can be causally attributed to the program) are about what-ifs

- Prospectively, we can think about how the world would be different *if we intervened* to change the status quo

- Retrospectively, we can think about what *would have been* had we not implemented a particular policy or program

- These alternative causal states are known as counterfactuals

- We are surrounded by what-ifs with potential relevance to population health—just look at the recent headlines...

**The New York Times**

Account ⌄

# *Taxing Farm Animals' Farts and Burps? Denmark Gives It a Try.*

Cows and pigs produce a large share of planet-warming methane. A new tax is part of a plan to clean that up.

Source: https://www.nytimes.com/2024/11/26/climate/denmark-methane-farm-animal-tax.html

## To fix broken mortgage markets, look to Denmark

Rising interest rates have exposed the problems with many home loans



IMAGE: GETTY IMAGES

Aug 31st 2023 · 3 min read

Fortunately there is a proven middle ground between the distorted American system and a floating-rate free-for-all. In Denmark homebuyers can borrow at 30-year fixed rates, and mortgages are prepayable. About half of borrowers fix for three decades. Yet there is no problem of "locked in" homeowners because a seller can end a mortgage by buying it back at its market value, which falls when rates rise, thereby cashing out the value of their interest-rate fix. Alternatively they can transfer their mortgage to the home's new owners. The result is greater dynamism: in the first quarter of 2023 housing transactions were down by only 6% on a year earlier, compared with 22% for existing homes in America.

There are catches: mortgage banks are protected from losses by creditor-friendly rules, including swift foreclosure procedures and barriers to declaring bankruptcy. But the flexibility and protection the Danish system offers is enviable when many mortgage markets are either distorting economies or causing hardship. Other countries would do well to try it. ■

Source: https://www.economist.com/leaders/2023/08/31/to-fix-broken-mortgage-markets-look-to-denmark

*The New York Times*

## Denmark Aims a Wrecking Ball at 'Non-Western' Neighborhoods

A government program is using demolition and relocation to remake neighborhoods with immigrants, poverty or crime.

Share full article    862

By **Emma Bubola**

Emma Bubola visited several neighborhoods around Denmark that the government defines as "parallel societies."
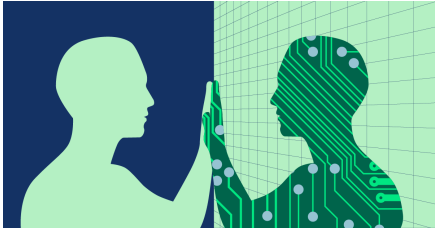
Oct. 26, 2023

After they fled Iran decades ago, Nasrin Bahrampour and her husband settled in a bright public housing apartment overlooking the university city of Aarhus, Denmark. They filled it with potted plants, family photographs and Persian carpets, and raised two children there.

Now they are being forced to leave their home under a government program that effectively mandates integration in certain low-income neighborhoods where many "non-Western" immigrants live.

In practice, that means thousands of apartments will be demolished, sold to private investors or replaced with new housing catering to wealthier (and often nonimmigrant) residents, to increase the social mix.

The Danish news media has called the program "the biggest social experiment of this century." Critics say it is "social policy with a bulldozer."

# Potential Outcomes Framework

- What-if or counterfactual questions are about hypotheticals—*so how can we answer them*?

- The potential outcomes framework provides us with a guide for posing and answering counterfactual questions; it is the common language for impact evaluation in the social sciences

- The potential outcomes framework uses the specification of well-defined causal states to which all members of the population of interest could be exposed to identify what **would have been** under an alternative counterfactual scenario

# Alternative treatment states

- Suppose that individuals $i$ in a population can be simultaneously assigned to two (or more) alternative treatments, $T_i$

- When only two alternative treatments are being considered they can be called the treatment and control states

Individual $i$

If "treated" $T_i = 1$

If "control" $T_i = 0$

James Lind is credited with introducing the concept of control and experimental groups; he is considered the father of clinical trials
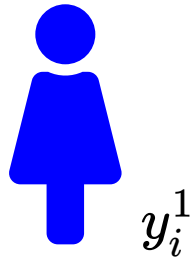
In 1747 Lind tested several scurvy treatments on crew members of the British navy and discovered that lemons and oranges were most effective.



Collier (2009)

# Potential Outcomes

- The potential outcomes for each individual are defined as the **true values** of the outcome that would result from exposure to well-defined, alternative causal states

- *Each individual has a potential outcome under both the treatment and control states*

- With a binary treatment, the potential outcomes are given by the random variables $Y^1$ and $Y^0$; we assume that each individual in the population has a potential outcome under both states

Potential outcomes $Y_i$ :

$y_i^1$

$y_i^0$

## Potential Outcomes

| | Treatment | Potential Outcomes | |
|---|---|---|---|
| **ID** | $d_i$ | $y_i^1$ | $y_i^0$ |
| 1 | 1 | 21 | 18 |
| 2 | 1 | 18 | 15 |
| 3 | 1 | 19 | 15 |
| 4 | 1 | 22 | 20 |
| 5 | 0 | 23 | 22 |
| 6 | 0 | 19 | 19 |
| 7 | 0 | 17 | 15 |
| 8 | 0 | 21 | 20 |

# The individual causal effect

Potential outcomes $Y_i$ :

$y_i^1$

$y_i^0$

Causal effect:

$$\delta_i = y_i^1 - y_i^0 \text{ or } \delta_i = y_i^1/y_i^0$$

- The **individual causal effect** is the difference in outcomes *for the same individual* with and without the intervention.

- A causal effect implies that the individual would have experienced the outcome if treated, but not untreated (or vice versa).
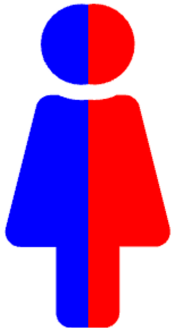
Individual causal effects

| | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| **ID** | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1 | 21 | 18 | 3 |
| 2 | 1 | 18 | 15 | 3 |
| 3 | 1 | 19 | 15 | 4 |
| 4 | 1 | 22 | 20 | 2 |
| 5 | 0 | 23 | 22 | 1 |
| 6 | 0 | 19 | 19 | 0 |
| 7 | 0 | 17 | 15 | 2 |
| 8 | 0 | 21 | 20 | 1 |

Individual causal effects

| | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| ID | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1 | 21 | 18 | 3 |
| 2 | 1 | 18 | 15 | 3 |
| 3 | 1 | 19 | 15 | 4 |
| 4 | 1 | 22 | 20 | 2 |
| 5 | 0 | 23 | 22 | 1 |
| 6 | 0 | 19 | 19 | 0 |
| 7 | 0 | 17 | 15 | 2 |
| 8 | 0 | 21 | 20 | 1 |

# A missing data problem

| Observed exposure | $Y^1$ | $Y^0$ |
|---|---|---|
| Treatment group | Observable as $Y$ | Counterfactual |
| Control group | Counterfactual | Observable as $Y$ |

- Although $Y_i^1$ and $Y_i^0$ exist in theory, in practice we cannot directly observe the same person simultaneously in two different states, which prevents direct calculation of individual causal effects
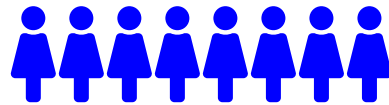
# The Average Treatment Effect (ATE)
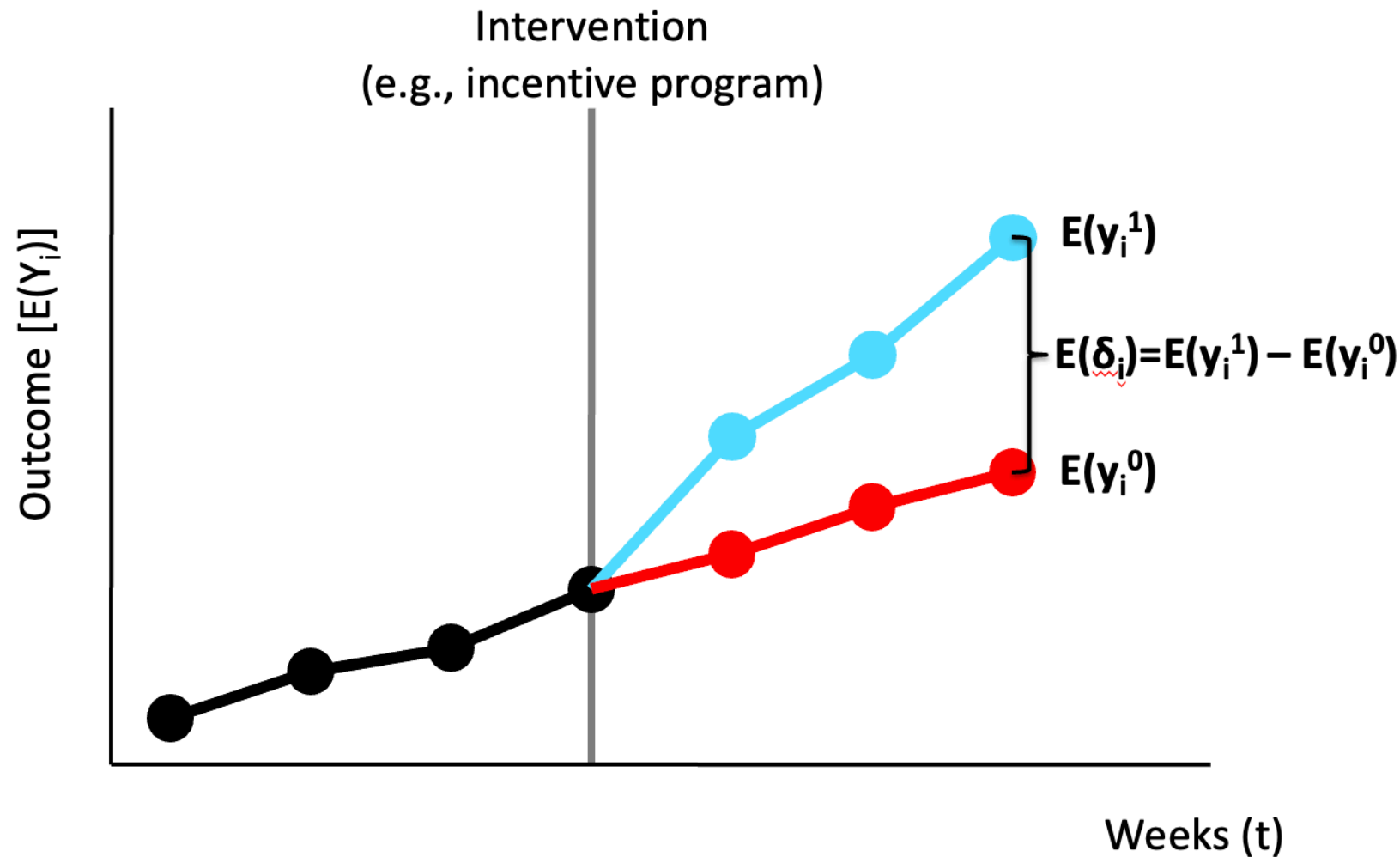
Our target population

Potential outcomes $E(Y_i)$:

$E(y_i^1)$

$E(y_i^0)$

- Causal effect is $\delta_i = y_i^1 - y_i^0$ or $\delta_i = y_i^1/y_i^0$

- The ATE compares potential outcomes *in the same population* under a treated scenario in which the policy is implemented and a control situation in which the policy is not implemented.
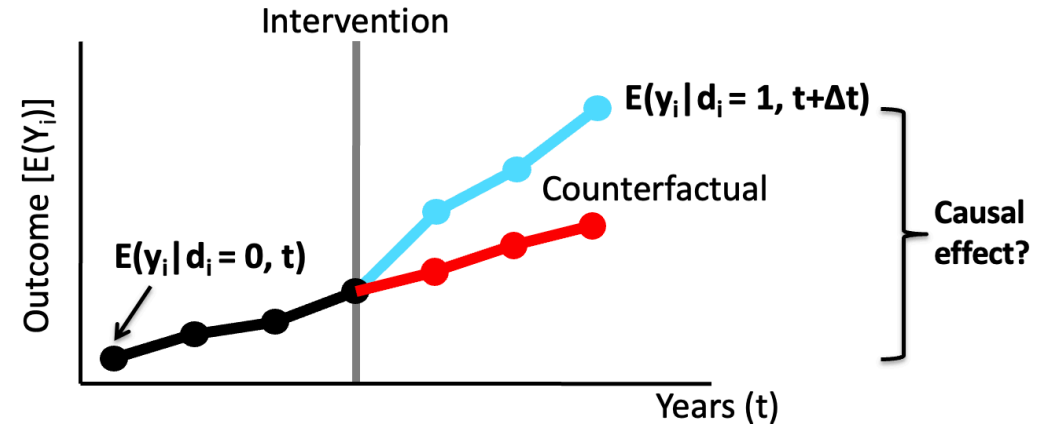
True (unobserved) impact of an intervention



Intervention
(e.g., incentive program)

Outcome [$E(Y_i)$]

$E(y_i^1)$

$E(\delta_i) = E(y_i^1) - E(y_i^0)$

$E(y_i^0)$

Weeks (t)

The average treatment effect

| ID | Treatment $d_i$ | Potential Outcomes $y_i^1$ | $y_i^0$ | Effect $\delta_i$ |
|---|---|---|---|---|
| 1 | 1.0 | 21 | 18 | 3 |
| 2 | 1.0 | 18 | 15 | 3 |
| 3 | 1.0 | 19 | 15 | 4 |
| 4 | 1.0 | 22 | 20 | 2 |
| 5 | 0.0 | 23 | 22 | 1 |
| 6 | 0.0 | 19 | 19 | 0 |
| 7 | 0.0 | 17 | 15 | 2 |
| 8 | 0.0 | 21 | 20 | 1 |
| **Mean** | **0.5** | **20** | **18** | **2** |

# Fundamental problem of causal inference

- As with individuals, it is not possible to observe the same target population simultaneously under two different conditions

- Unlike their individual-level analogues, we can use our observed data to estimate $E(y_i)$ and calculate a "naïve" estimate of the ATE
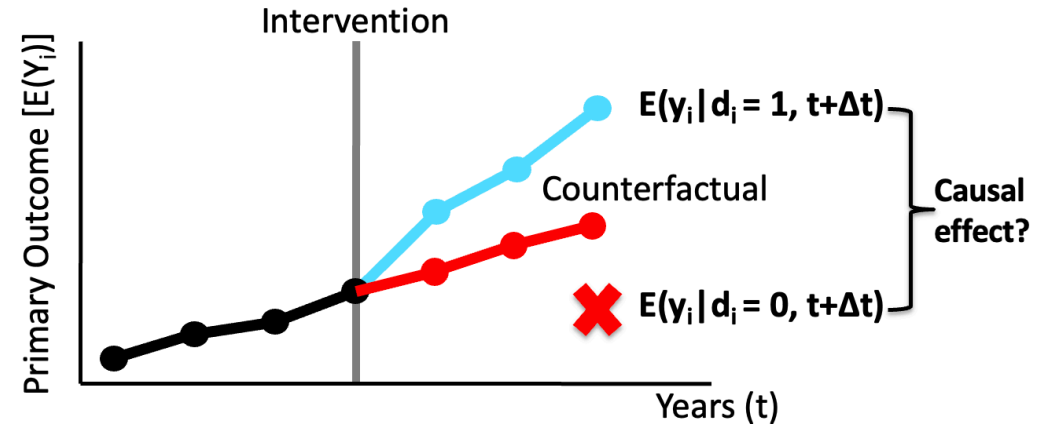
# Fundamental problem of causal inference

- As with individuals, it is not possible to observe the same target population simultaneously under two different conditions

- Unlike their individual-level analogues, we can use our observed data to estimate $E(y_i)$ and calculate a "naïve" estimate of the ATE



Intervention

Outcome [$E(Y_i)$]

$E(y_i | d_i = 1, t+\Delta t)$

Counterfactual

Causal effect?

$E(y_i | d_i = 0, t)$

Years (t)

- We could observe the same group at different time periods (pre-post), but other things may have changed since the intervention
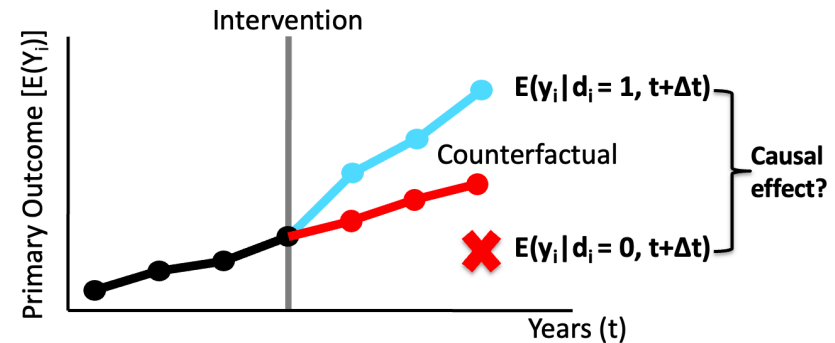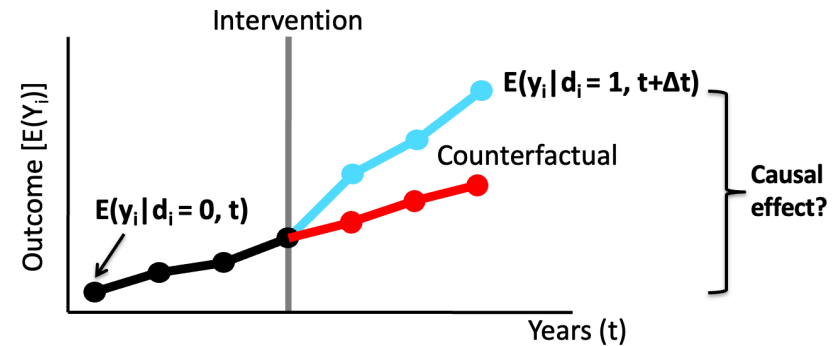
# Fundamental problem of causal inference

- As with individuals, it is not possible to observe the same target population simultaneously under two different conditions

- Unlike their individual-level analogues, we can use our observed data to estimate $E(y_i)$ and calculate a "naïve" estimate of the ATE



- Alternatively, we can observe a different group that was unexposed

# These substitutes are not good counterfactuals

- Pre-post ignores "secular trends" (changes in other factors influencing the outcome since the intervention)

- Treated vs. control neglects unmeasured differences between the groups

# Naive estimate of the ATE

Population showing actual treatment status, $D_i$
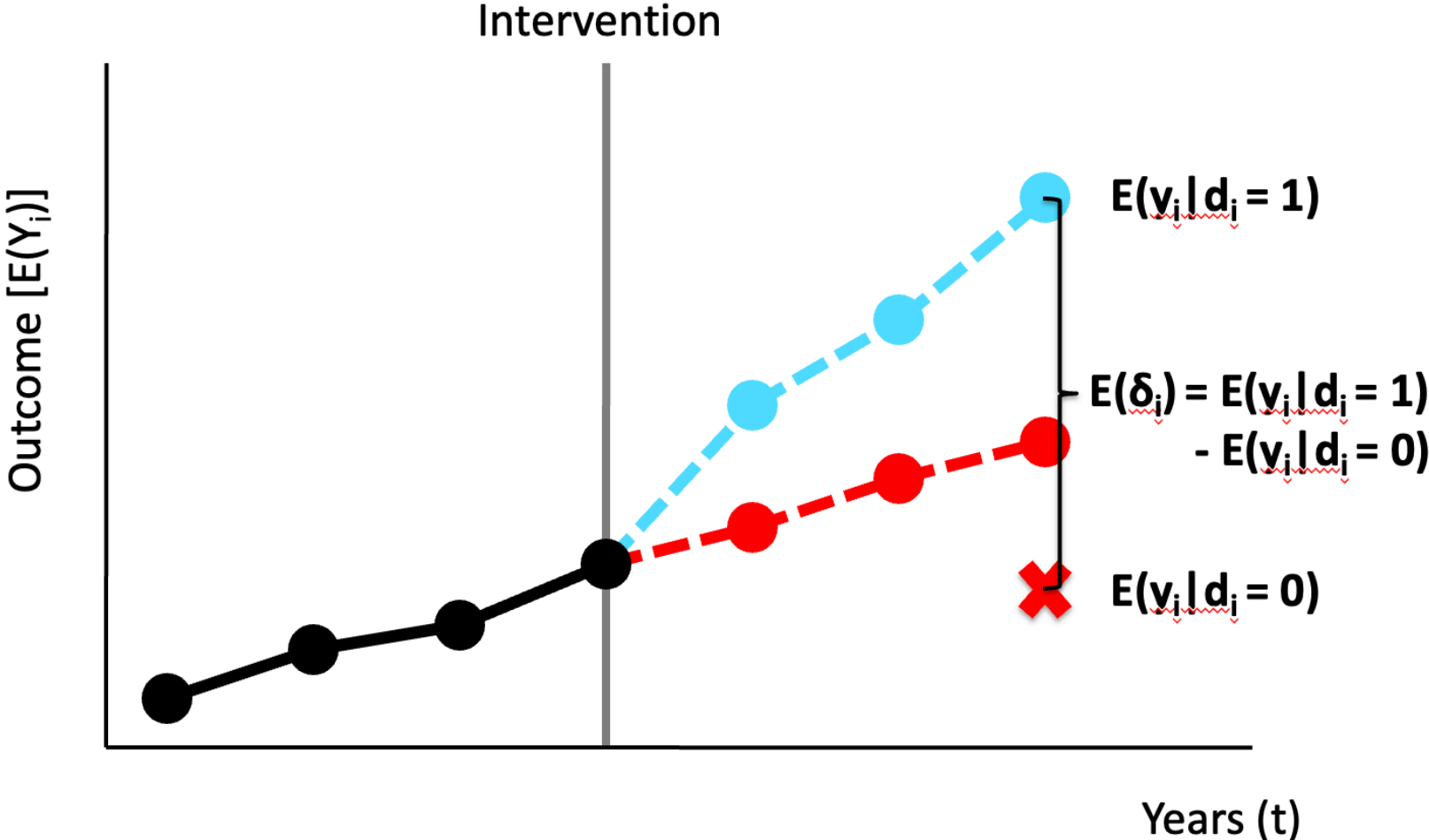
Observed
outcomes
$E(Y_i|D_i)$:

$$E(y_i|d_i = 1)$$

$$E(y_i|d_i = 0)$$

- Naive estimate is: $E(\delta_{i,naive}) = E(y_i|d_i = 1) - E(y_i|d_i = 0)$

- The naïve estimate of the ATE compares the outcomes for those observed to receive the treatment vs. the control condition

# Estimated (observed) impact of the program



Intervention

Outcome [$E(Y_i)$]

$E(v_i | d_i = 1)$

$E(\delta_i) = E(v_i | d_i = 1)$
$- E(v_i | d_i = 0)$

$E(v_i | d_i = 0)$

Years (t)

# The selection problem

The naïve estimator rarely yields an unbiased treatment effect

Why? Because treatment and control groups are rarely exchangeable (i.e., programs/policies are selectively placed in different areas and the decision to participate is often voluntary)

These differences could affect potential outcomes, creating bias

Economist call this selection or omitted variable bias

In epidemiology, the effect of these pre-existing differences between groups is commonly called confounding bias

# Recovering causal effects

Without a credible substitute for the unobserved potential outcomes, causal impact cannot be established.

The key is to generate groups of individuals that are statistically indistinguishable from one another in the absence of an intervention —we must somehow mimic the counterfactual
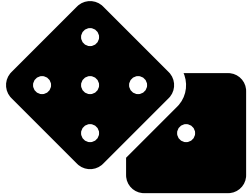
We can use experimental or quasi-experimental techniques to attempt to recover the causal effect

# What are randomized experiments?

> the word experiment is used in a quite precise sense to mean an investigation where the system under study is under the control of the investigator. This means that the individuals or material investigated, the nature of the treatments or manipulations under study and the measurement procedures used are all selected, in their important features at least, by the investigator. (Cox, 2000)

Randomized experiments are experiments "in which units are assigned to receive the treatment or an alternative condition by a random process such as the toss of a coin." (Shadish, 2002)

# RCTs and exchangeability

If treatment is assigned randomly, it is said to be exogenous

Comparison groups are the same with respect to measured and unmeasured characteristics (no unmeasured confounding or omitted variable bias) and each unit's potential outcomes are independent of actual treatment assignment, $(Y_i^1, Y_i^0 | D)$

Under these conditions, the treatment and control groups are exchangeable, which implies that **the probability of the outcome in the control group is the same as the probability of the outcome in the treated group had they been untreated** (and vice versa)

Randomization is so highly valued because, under ideal conditions, it provides an unbiased estimate of the causal impact of the intervention in the population under study
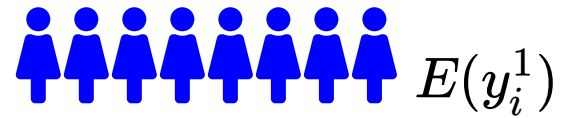
Target population
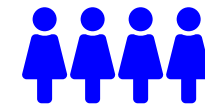
Randomization

Actual exposure

Causation

$E(y_i^1)$

vs

$E(y_i^0)$

$=$

Association

$E(y_i|d_i = 1)$

vs

$E(y_i|d_i = 0)$

# 'Quasi'-experiments

Oftentimes we do not have control over treatment assignment and must investigate how individuals end up in alternative states

The term quasi-experiment refers to:

> "experiments that have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred."(Cook, 1979)

Assignment is by self-selection, by which units choose treatment for themselves, or by means of administrator selection

Consequently, unlike RCTs, comparison groups may differ in many systematic ways other than the presence of the treatment

# Causal Policy Effects

I. "What-If" Questions and Counterfactuals

**II. Causal Parameters**

III. Causal Assumptions

# Treatment effects

**?**

So far we have discussed the average treatment effect or ATE

The ATE might, however, not be the only quantity of interest

The causal parameters we can estimate will depend on our research question, and the design of our evaluation study

# Average Treatment Effect on the Treated (ATT)

It might make sense to evaluate the impact for those who self-select into certain exposures/programs, in which case a conditional average treatment effect may be more relevant; the ATT is the most widely-used measure of average impact

The ATT is the average treatment effect for those who typically take the treatment:

$$E(\delta_i|d_i = 1) = E(y_i^1|d_i = 1) - E(y_i^0|d_i = 1)$$

Unbiased estimation requires:

$$E(y_i^0|d_i = 1) = E(y_i^0|d_i = 0)$$

Thus, it relies on identification of one counterfactual, $E(y_i^0|d_i = 1)$

Calculate the ATE and the ATT

| | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| ID | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1.0 | 21 | 18 | 3 |
| 2 | 1.0 | 18 | 15 | 3 |
| 3 | 1.0 | 19 | 15 | 4 |
| 4 | 1.0 | 22 | 20 | 2 |
| **Mean** | **1.0** | **20** | **17** | **3** |
| 5 | 0.0 | 23 | 22 | 1 |
| 6 | 0.0 | 19 | 19 | 0 |
| 7 | 0.0 | 17 | 15 | 2 |
| 8 | 0.0 | 21 | 20 | 1 |
| **Mean** | **0.0** | **20** | **19** | **1** |
| **Mean** | **0.5** | **20** | **18** | **2** |

# Calculate the ATE

| | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| **ID** | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1.0 | 21 | 18 | 3 |
| 2 | 1.0 | 18 | 15 | 3 |
| 3 | 1.0 | 19 | 15 | 4 |
| 4 | 1.0 | 22 | 20 | 2 |
| **Mean** | **1.0** | **20** | **17** | **3** |
| 5 | 0.0 | 23 | 22 | 1 |
| 6 | 0.0 | 19 | 19 | 0 |
| 7 | 0.0 | 17 | 15 | 2 |
| 8 | 0.0 | 21 | 20 | 1 |
| **Mean** | **0.0** | **20** | **19** | **1** |
| **Mean** | **0.5** | **20** | **18** | **2** |

## Calculate the ATT

|  | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| ID | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1.0 | 21 | 18 | 3 |
| 2 | 1.0 | 18 | 15 | 3 |
| 3 | 1.0 | 19 | 15 | 4 |
| 4 | 1.0 | 22 | 20 | 2 |
| **Mean** | **1.0** | **20** | **17** | **3** |
| 5 | 0.0 | 23 | 22 | 1 |
| 6 | 0.0 | 19 | 19 | 0 |
| 7 | 0.0 | 17 | 15 | 2 |
| 8 | 0.0 | 21 | 20 | 1 |
| **Mean** | **0.0** | **20** | **19** | **1** |
| **Mean** | **0.5** | **20** | **18** | **2** |

# More on the ATE vs. the ATT

Example from a basic income program for the unemployed

- The ATE is the expected what-if difference in the outcome if we could take *randomly selected* unemployed workers and compare their outcomes in both the program vs. not

- The ATT is the expected what-if difference in the outcome if we could observe workers *enrolled* in the basic income program under two scenarios, a factual one in which they were enrolled and the other in which they were not

- Thus, the ATT answers the question about whether the social program was beneficial specifically among those who were enrolled, whereas the ATE examines whether it was beneficial across all of the unemployed workers who could potentially enroll, including those who did and those who didn't

# Average Treatment Effect on the Controls (ATC)

A less common contrast is given by the average treatment effect on the controls (ATC):

$$E(\delta_i|d_i = 0) = E(y_i^1|d_i = 0) - E(y_i^0|d_i = 0)$$

The ATC answers the questions about whether the program is beneficial for those who are not enrolled (e.g., would a program that attempts to enroll nonparticipants be beneficial?

Unbiased estimation requires:

$$E(y_i^1|d_i = 0) = E(y_i^1|d_i = 1)$$

Thus, it relies on identification of one counterfactual, $E(y_i^1|d_i = 0)$

# Calculate the ATC

| | Treatment | Potential Outcomes | | Effect |
|---|---|---|---|---|
| **ID** | $d_i$ | $y_i^1$ | $y_i^0$ | $\delta_i$ |
| 1 | 1.0 | 21 | 18 | 3 |
| 2 | 1.0 | 18 | 15 | 3 |
| 3 | 1.0 | 19 | 15 | 4 |
| 4 | 1.0 | 22 | 20 | 2 |
| **Mean** | **1.0** | **20** | **17** | **3** |
| 5 | 0.0 | 23 | 22 | 1 |
| 6 | 0.0 | 19 | 19 | 0 |
| 7 | 0.0 | 17 | 15 | 2 |
| 8 | 0.0 | 21 | 20 | 1 |
| **Mean** | **0.0** | **20** | **19** | **1** |
| **Mean** | **0.5** | **20** | **18** | **2** |

# Relation between ATE, ATT, and ATC

- The ATE is the average of the ATT and ATC, weighted by the proportions who were treated ($\pi$) or control ($1 - \pi$), respectively:

$$ATE = \pi * E(\delta|D = 1) + (1 - \pi) * E(\delta|D = 0)$$

- In a RCT, the treatment effects should be the same among those randomized to treatment vs. control (ATE=ATT=ATC).

- Why?

# Simple example

Using the table and assuming $\pi = 0.3$, calculate the:

| Group | $E[Y^1|.]$ | $E[Y^0|.]$ |
|---|---|---|
| Treatment group $(D = 1)$ | 10 | 6 |
| Control group $(D = 0)$ | 8 | 5 |

- naïve estimate of the ATE,
- the ATE
- the ATT
- the ATC

# Causal Policy Effects

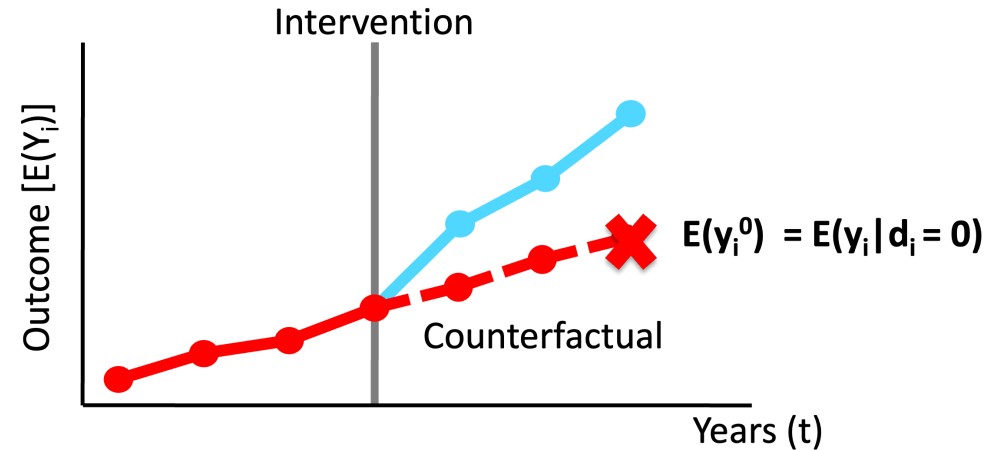I. "What-If" Questions and Counterfactuals

II. Causal Parameters
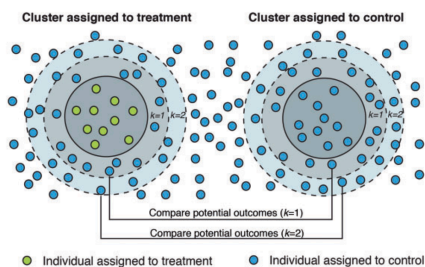
**III. Causal Assumptions**

# Exchangeability

For the observed contrasts to represent causal effects, we need the proxies to represent the unobserved counterfactual outcomes

If the average outcome in the control group equals the average outcome in the treated group had it been untreated (and vice versa) , the groups are exchangeable, i.e., for all $d$, $E(Y_i^d) = E(Y_i|D = d)$

# Stable Unit Treatment Value Assumption (SUTVA)



Cluster assigned to treatment    Cluster assigned to control

Compare potential outcomes (k=1)
Compare potential outcomes (k=2)

○ Individual assigned to treatment    ● Individual assigned to control

The Stable Unit Treatment Value Assumption (SUTVA) assumes that an individual's potential outcome is unaffected by the treatment status of other individuals

This assumption is most commonly violated when there is interference (an individual's outcome depends on the treatment values of other individuals), or in the case of ecological treatments, if there are spillover effects

Say we wanted to evaluate the impact of an education intervention assigned to randomly selected children in classrooms on substance use—what are some potential violations of SUTVA?
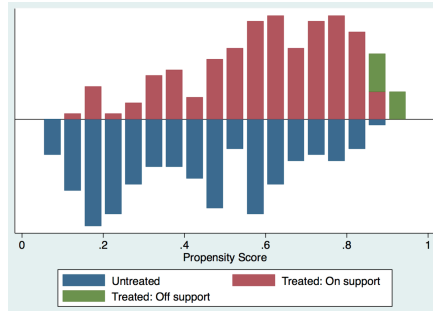
# Consistency

$Y_j^{obs} = Y_j(x)$ if
$x = X_j$

> individual $j$'s potential outcome $Y_j(x)$ under exposure $x$ as the outcome that would have been observed if individual $j$ had received exposure $x$

Causal inference requires that a treatment is defined unambiguously, which a general concern in social epidemiology

This vagueness of some potential treatments has led some to suggest that only the causal effects of treatments that can be hypothetically manipulated should be considered (i.e., Holland's "no causation without manipulation")

However, consistency violations are less of a concern in impact evaluation (and guaranteed by experiments) since we are dealing with treatments that are, by definition, manipulable

See Cole and Frangakis (2009)

# Positivity



The positivity assumption states that there is a nonzero (ie, positive) probability of receiving every level of exposure for every combination of values of exposure and confounders that occur among individuals in the population

Generally less of a concern for RCTs, non-positivity, sometimes called structural confounding, is a concern for observational designs that often rely on regression adjustment for confounders to achieve conditional exchangeability

# Further reading

Holland, Paul W. "Statistics and causal inference." Journal of the American statistical Association 81.396 (1986): 945-960.

Morgan, Stephen L., and Christopher Winship. Counterfactuals and causal inference, Chapters 1 and 2. Cambridge University Press, 2007.

William R. Shadish, Thomas D. Cook, and Donald Thomas Campbell. Experimental and quasi-experimental designs for generalized causal inference, Chapter 1. Wadsworth Cengage learning, 2002.

Winship, Christopher, and Stephen L. Morgan. "The estimation of causal effects from observational data." Annual review of sociology (1999): 659-706.

# Break! ☕

15:00