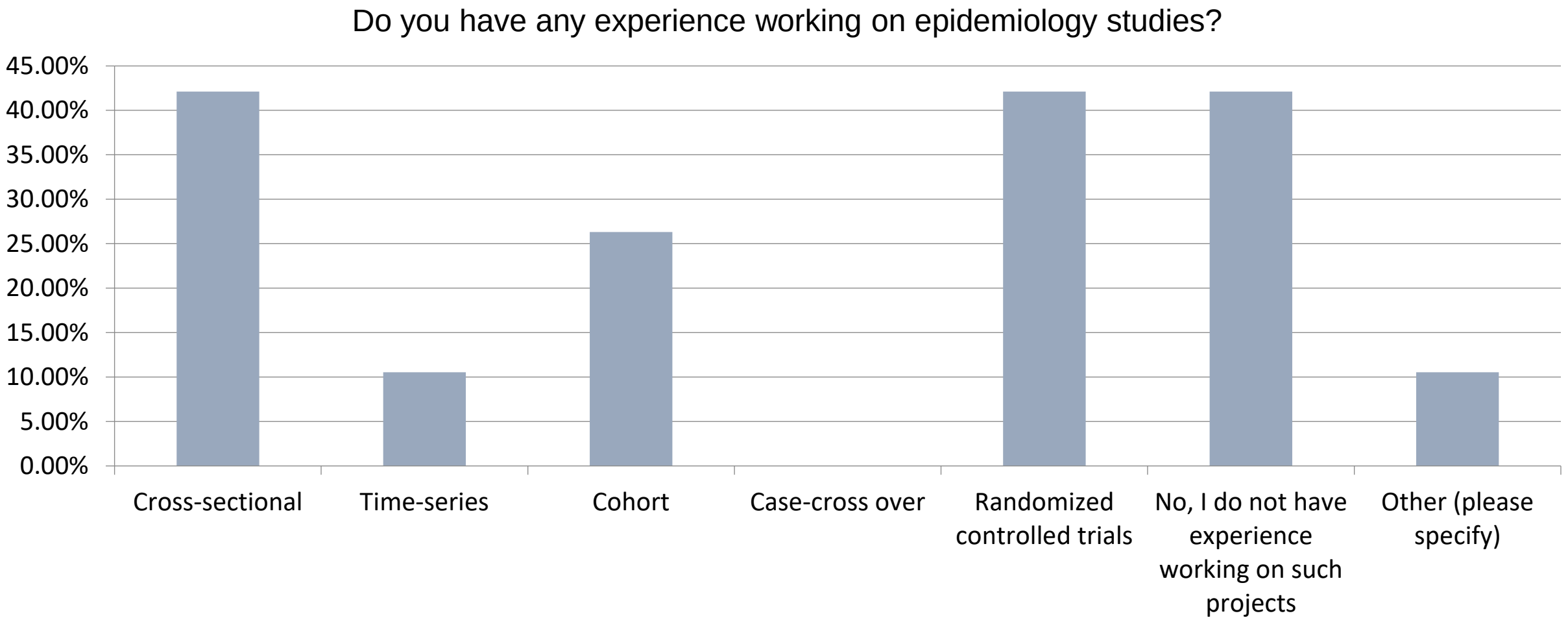
Introduction to air pollution epidemiology

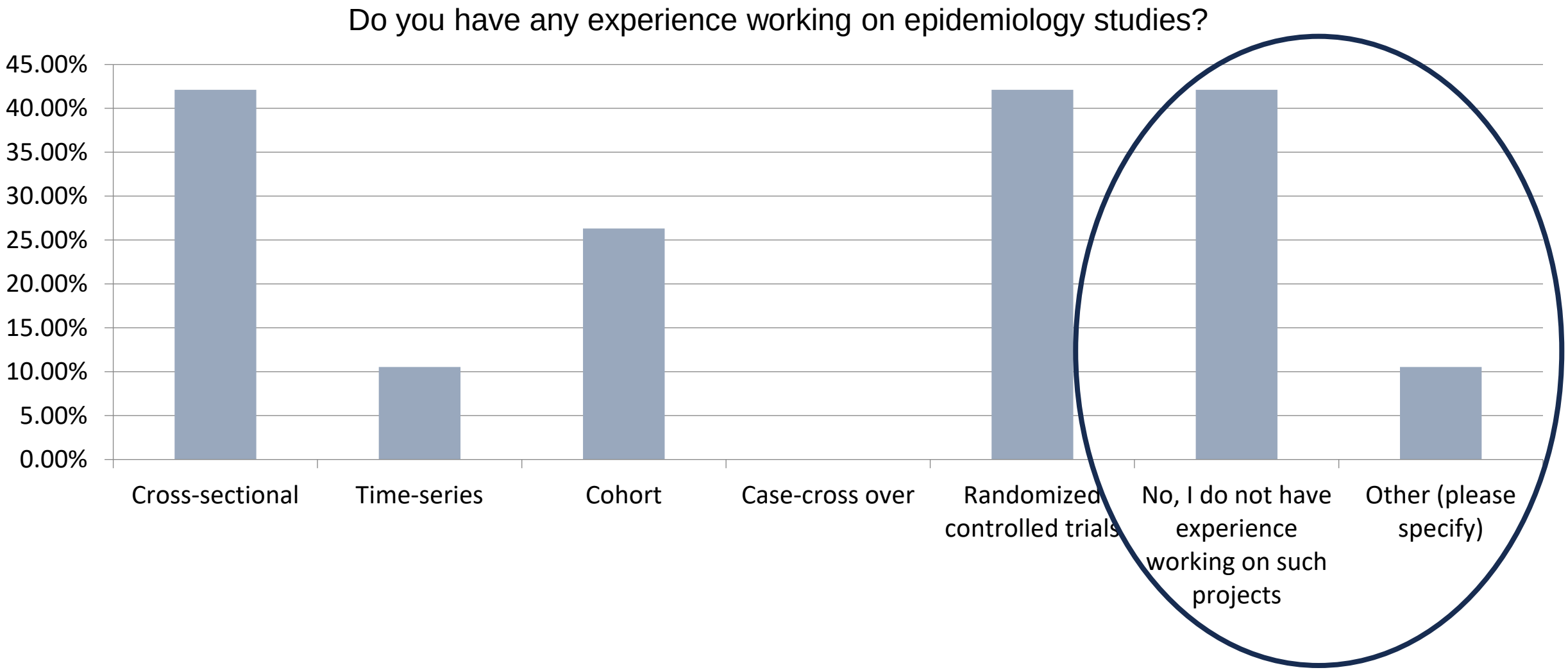
Jill Baumgartner

Professor

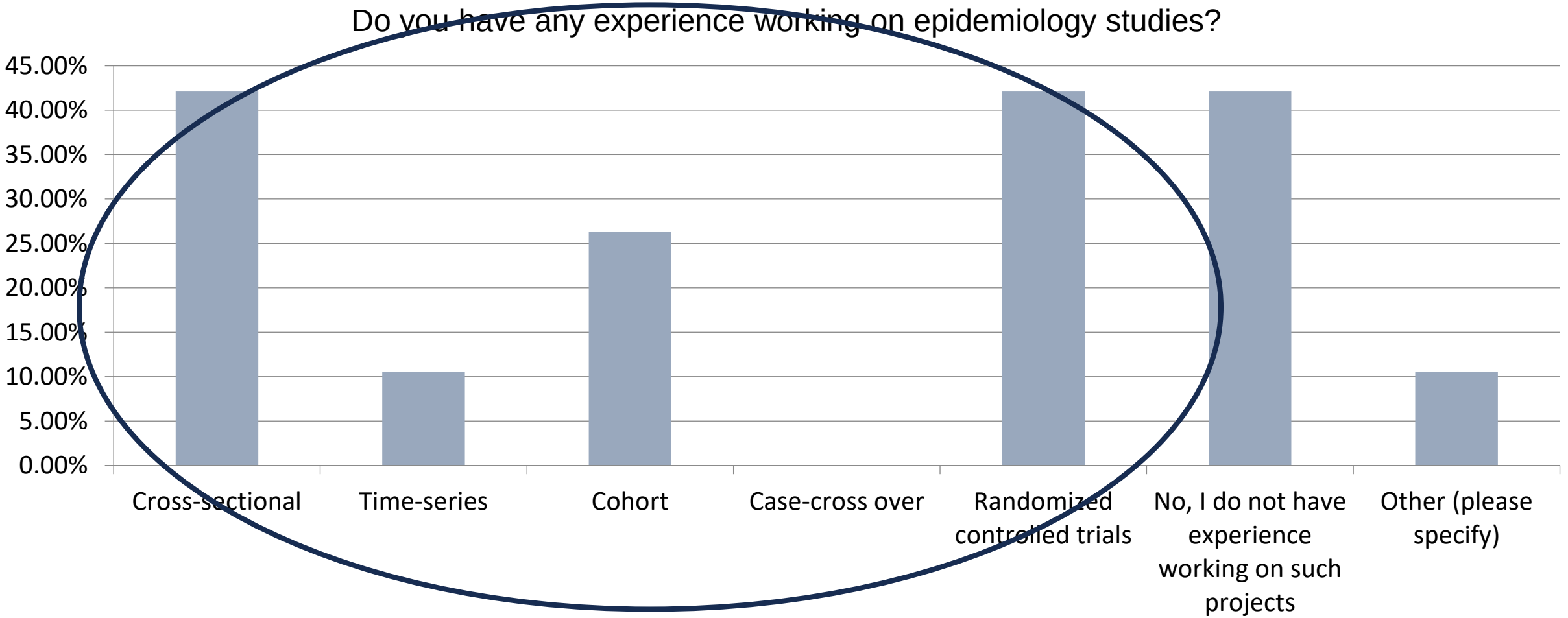
School of Population and Global health

McGill University, Canada


Do you have any experience working on epidemiology studies?
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
Cross-sectional
Time-series
Cohort
Case-cross over
Randomized controlled trials
No, I do not have experience working on such projects
Other (please specify)



Learning objective 1:
Understand the goals of study design in air pollution epidemiology – what are we trying to accomplish



Learning objective 2:

Be familiar with the main epidemiologic designs used to evaluate the health impacts of air pollution or related interventions and policies

There is no such thing as a perfect study design.

It is our responsibility as researchers with integrity to understand and acknowledge the limitations in our research, and to be transparent about how those limitations might affect our results and conclusions.

Limitations are not something to hide.

What is air pollution epidemiology?

- Air pollution is one of the most widely studied environmental exposures - ubiquitous, modifiable, and affects a range of health outcomes throughout the lifecourse

- A subset of epidemiology, which is is the **study** of the **distribution** and **determinants** of health in **populations** *(Last, 2001)*

- Focuses on air pollution or air pollution-related interventions as determinants of health outcomes

What is the goal of air pollution epidemiology?

What is the goal of air pollution epidemiology?

Provide evidence for establishing causal effects of (a) air pollution on health outcomes and (b) the health impacts of air pollution interventions or policies

the process of inferring whether an observed correlation reflects a causal relationship based on evidence and reasoning.

The counterfactual model

David Hume

Scottish philosopher (1711 – 1776)

"…we may define a cause to be an object followed by another…where, if the first object had not been, the second never had existed."

Hume's definition was the first iteration of what we call the counterfactual condition: a hypothesis about what would have happened under conditions contrary to fact (i.e., counter-factual).

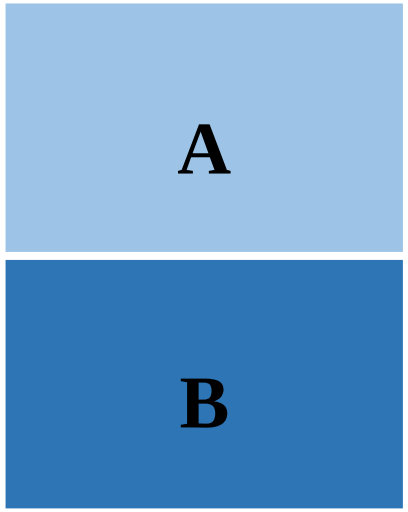
Average causal effect in a group of individuals (i.e., a population)

For a specified target population and time period, the average risk is:

$$R = \frac{A}{B}$$

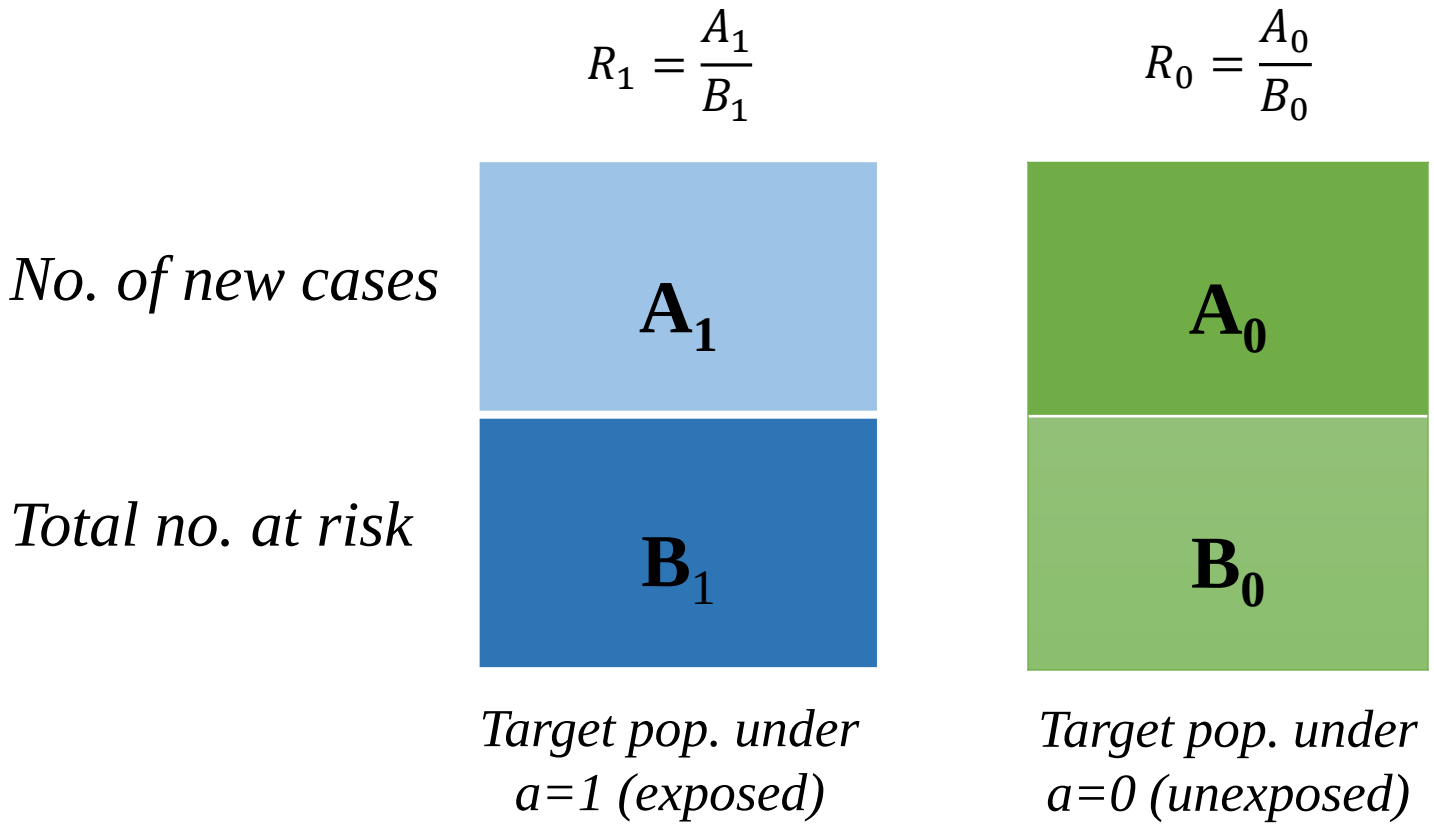
No. of new cases

Total no. at risk



Target population

Under the counterfactual model, to estimate the causal effect of an air pollution we need to observe the risk in the *same target population* under different exposure conditions *at the same time*.



If R_1 is the average risk had the target population been exposed and R_0 is the average risk had the target population not been exposed, we can compare those risks:

$$Risk\ ratio_{causal} = \frac{R_1}{R_0}$$

$$Risk\ difference_{causal} = R_1 - R_0$$

These quantities, risk ratios and risk differences,

are known as effect measures or causal contrasts.

Average causal effect

- Causal effect measures compare measures of disease occurrence, R_1 and R_0, under two different exposure conditions *in the same target population at the same time*.

- We can measure the causal effect in this scenario because the risk of the outcome in the exposed group is equal to the risk in the unexposed group if it had been exposed instead.

- The two groups are *exchangeable.*
  - The concept of exchangeability is crucial to every possible research design.

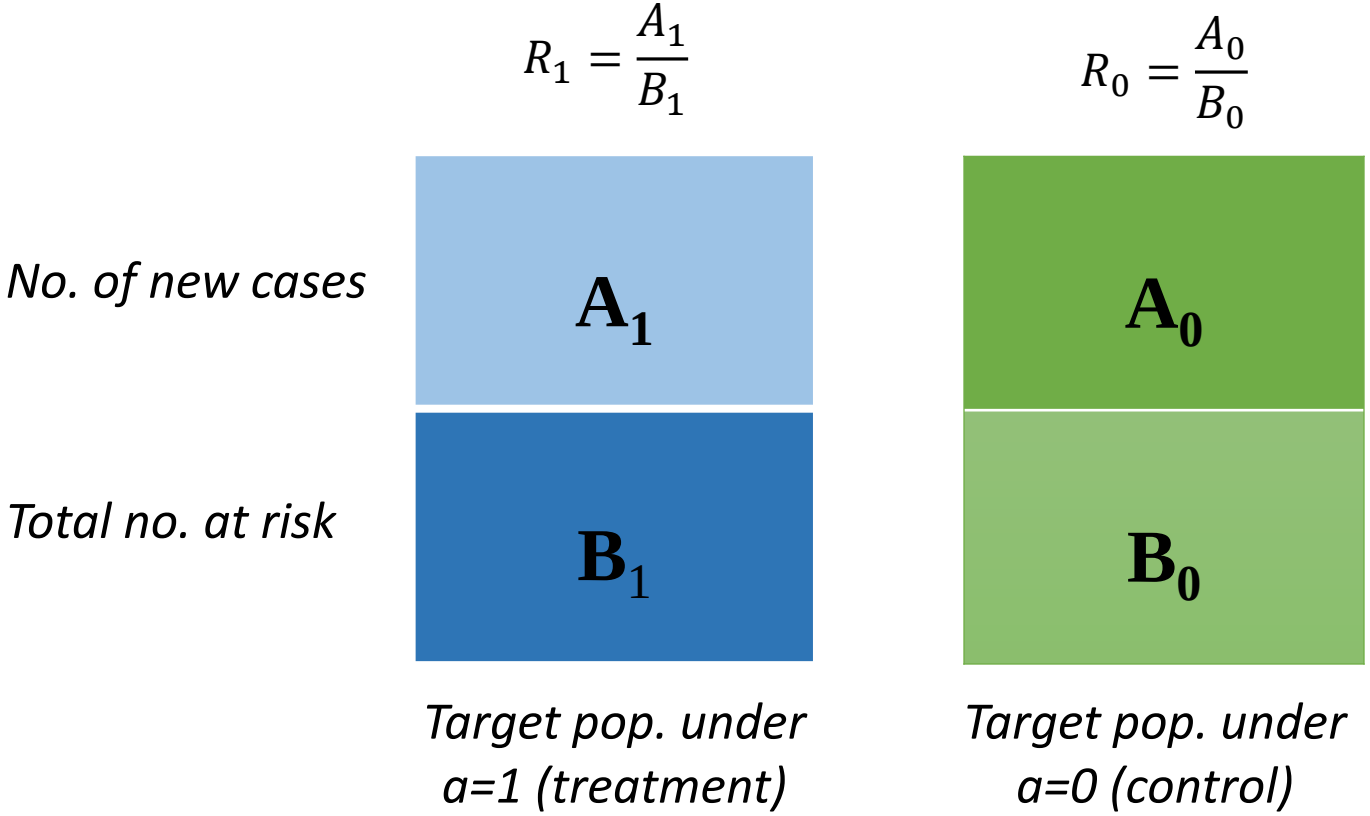
- Any difference between R_1 and R_0 is attributable to the exposure.

Obviously, we cannot observe the same population under two different exposures at the same time.

At least one of the two exposure conditions is hypothetical, or counter-to-fact.

This is the fundamental problem of causal inference and makes measuring causal effects *directly* impossible.

Under the counterfactual framework, we know that our ultimate and unachievable goal is to measure the causal effect of exposure on outcome by simultaneously observing the same target population under different exposure conditions.



Exchangeability

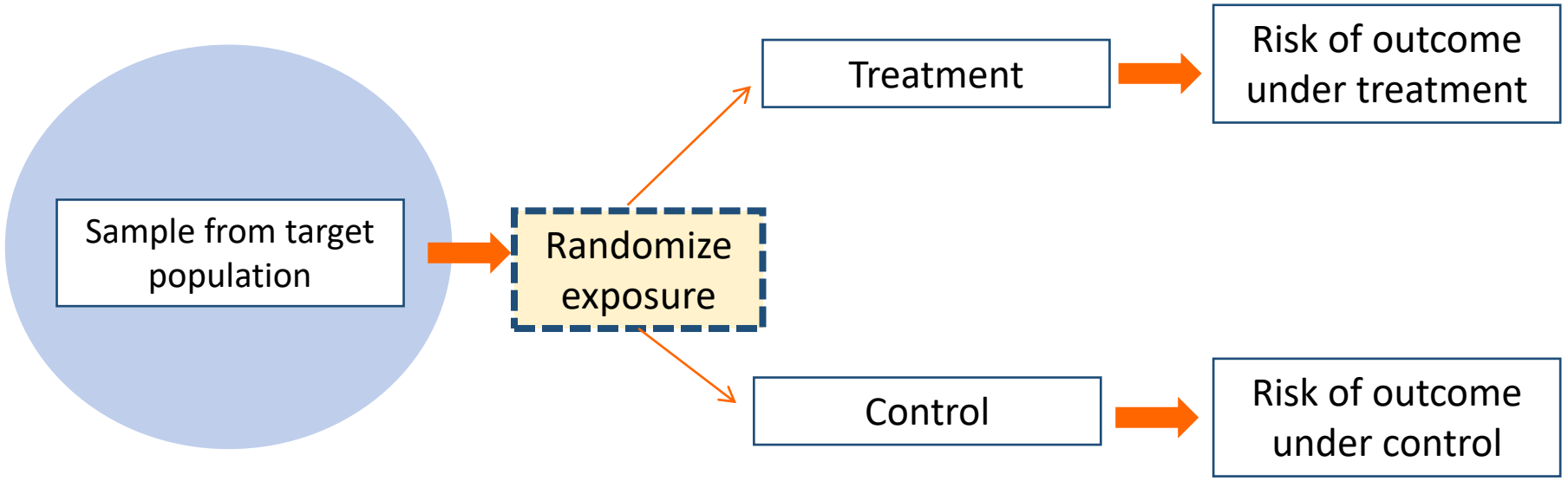
- If we could observe the same population simultaneously under different exposure conditions, we would have perfect exchangeability.

- The risk of the outcome in the exposed group (R_1) would equal the risk of the outcome in the unexposed group (R_0) *had it been treated*.

- Under perfect exchangeability, any difference between R_1 and R_0 is attributable to the exposure.

Randomized trials

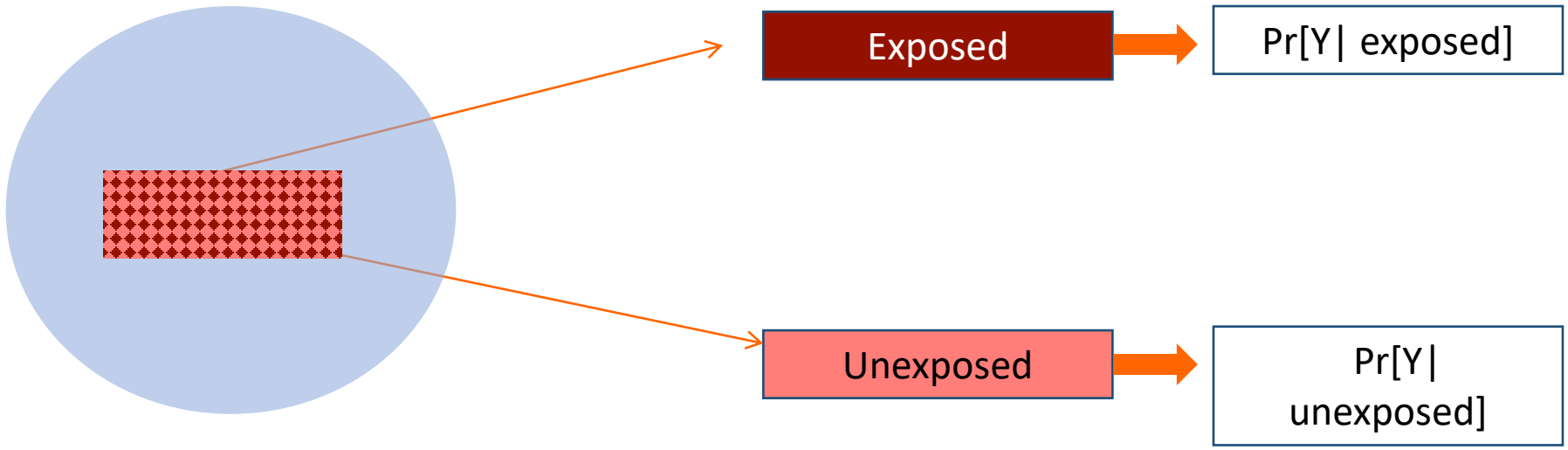
We know that randomization represents our best effort to achieve exchangeability in the real world.



When randomization is successful, the groups being compared are exchangeable with regard to both measured and unmeasured characteristics.

Observational studies

Exposure is not randomized in air pollution epidemiology (observational) studies and so there is no reason to believe that the groups we are comparing are exchangeable.



Persons who are exposed and persons who are unexposed may differ in many ways *in addition to their exposure status*.

Substitution

When *estimating* causal effects in the real world, we are forced to calculate either R_1 or R_0 based on a substitute population.

The validity of our estimates of risk depend upon the validity of the substitution.

Measures of *association*

There is no reason to expect that the observed outcome in the substitute population is equivalent to the outcome we would have observed in the target population under the same exposure (a=0 or a=1).

Therefore, we must substitute a **measure of association** for the causal contrast or causal effect

$$RR_{association} = \frac{R_1}{Substitute\ for\ R_0} = \frac{A_1/B_1}{E_0/F_0}$$

$$RD_{association} = R_1 - Substitute\ for\ R_0 = \left(\frac{A_1}{B_1}\right) - \left(\frac{E_0}{F_0}\right)$$

What assumptions are being made?

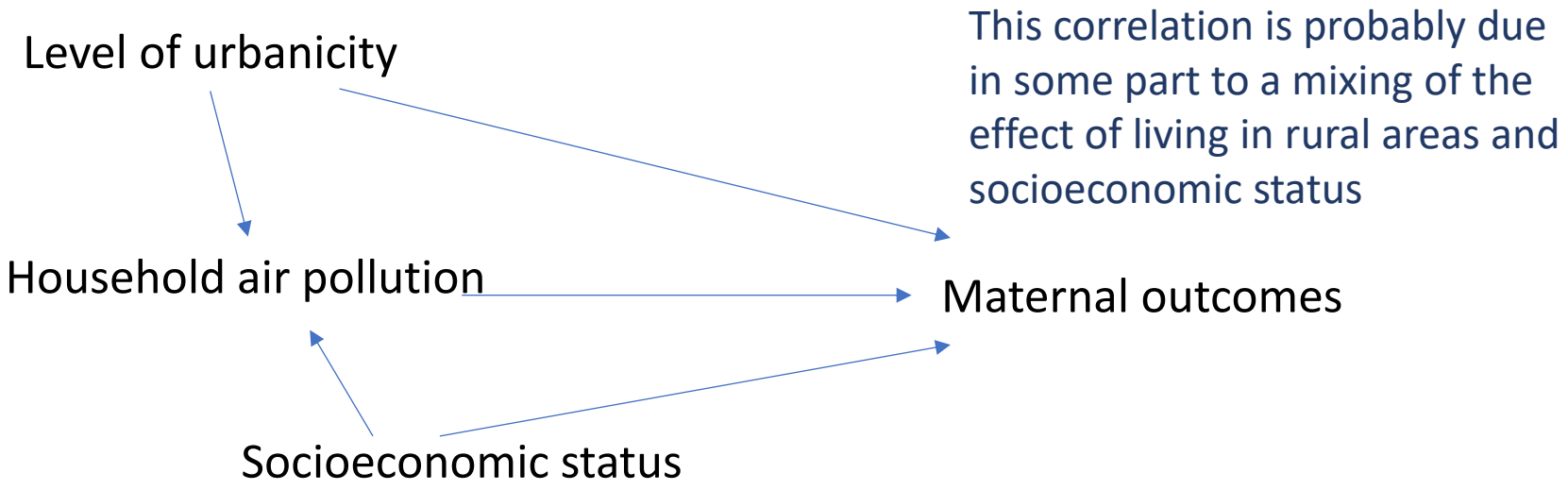
Confounding is a lack of exchangeability

Without exchangeability, any differences between R_1 and R_0 could be wholly or partially attributable to the exposure we're interested in…or wholly or partially attributable to other differences between the groups being compared.

This lack of exchangeability gives rise to confounding.

Confounding is a mixing of effects

There is a strong correlation between household air pollution and socioeconomic status.



This correlation is probably due in some part to a mixing of the effect of living in rural areas and socioeconomic status

Mothers exposed to household air pollution may be very different from mothers who are not.

THEY ARE NOT EXCHANGEABLE.

The counterfactual model should inform the design of your air pollution-health study

The counterfactual model provides a framework for designing and analyzing etiologic studies of air pollution and health.

All studies should be designed to estimate causal contrasts (this is the goal of air pollution epidemiology!), and the specific contrast of interest should be carefully defined in advance.

1. Define a target population and time period.
2. Clearly define the causal contrast of interest.
3. Find substitutes for counterfactual measures.

For further reading, see Maldonado G and Greenland S. Estimating causal effects. *International Journal of Epidemiology*. 2002;31:422-429.

Air pollution epidemiology: short- versus long-term study designs

- Air pollution-health studies often categorized in two broad categories
  - Short-term: evaluate acute impacts of exposures over hours, days, or weeks
  - Long-term: assess chronic impacts of exposures over months, years, decades, or lifetimes

- Health outcomes of short- & long-term exposures can be different:
  - Short term: respiratory symptoms, asthma exacerbation, blood pressure, hospitalizations, deaths
  - Long-term: development of chronic conditions (asthma, COPD, cardiovascular disease), premature death

Designs for studying short-term health effects of air pollution

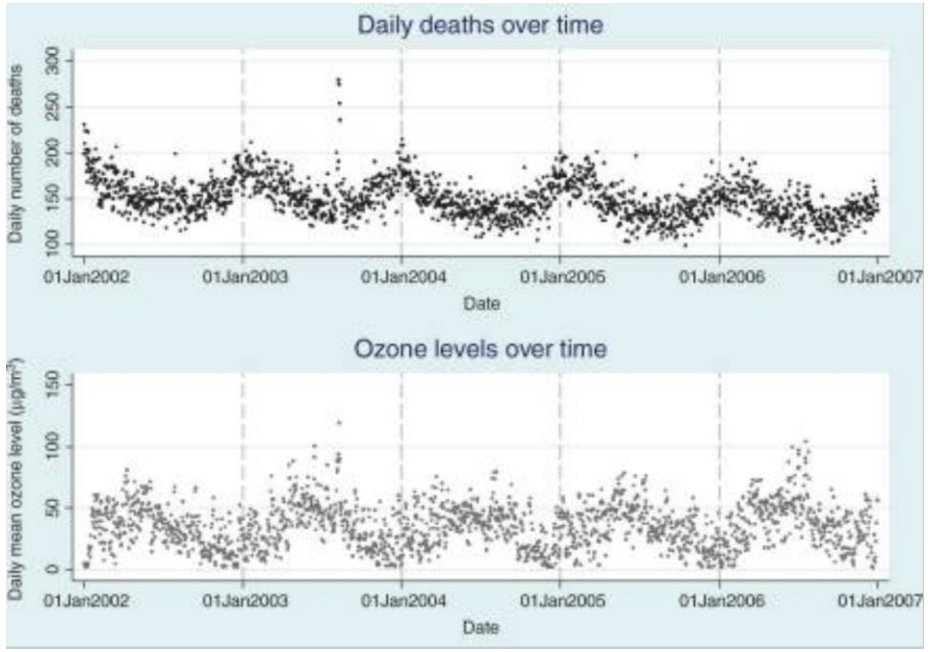
- Time-series studies*
- Case-crossover studies*

*there are many iterations of these designs at population and individual levels

Time-series studies

- Repeated (daily) measures of exposure and outcomes over time
- Associate daily variations in exposure with daily variation in health (e.g., hospitalizations, deaths)
- Can account for temporal trends over time
- Must adjust for time-varying confounders (e.g., temperature, pollen, day of the week, holidays, seasonal influenza)

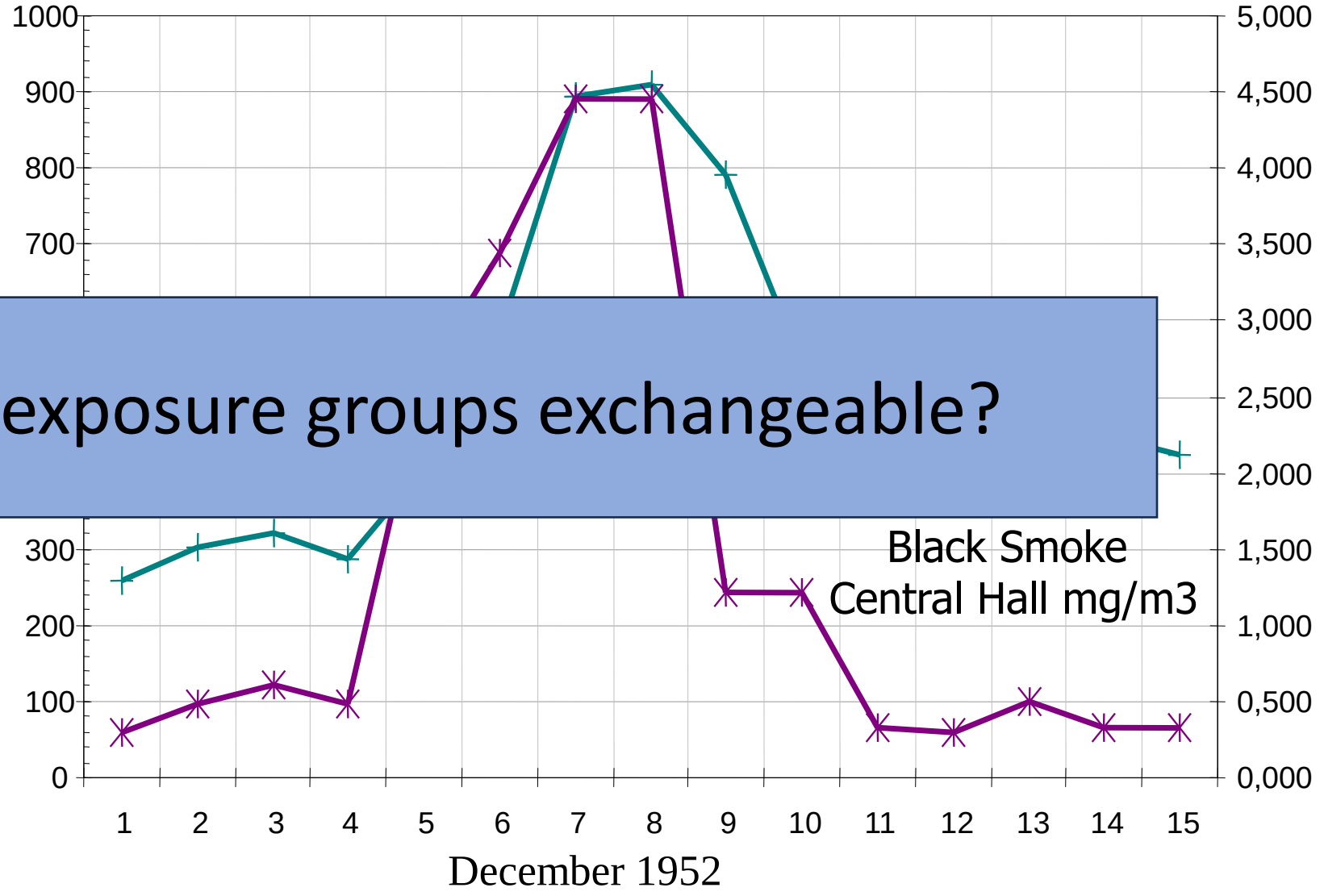
Scatterplots of daily ambient ozone number of deaths over time in London



For an instructive example, see Bhaskaran et al., Int J Epi, 2013 (doi: 10.1093/ije/dyt092)

Great London Smog (1952)

- Example of a time-series study that assessed daily air pollution and daily m
- D
  D
  ti
  over comparable periods in 1947-1951



Are these exposure groups exchangeable?

Case-crossover study

- Also used to analyze time-series data but compares individuals to themselves
- Common design is to compare and individual's air pollution exposure immediately prior to the health event (index time) to the same person's exposure at otherwise similar reference times (control periods)

For an overview of different self-matched methods, see Mostofsky E, Coull BA, Mittleman MA. Analysis of observational self-matched data to examine acute triggers of outcome events with abrupt onset. Epidemiology. 2018;29(6):804–816.

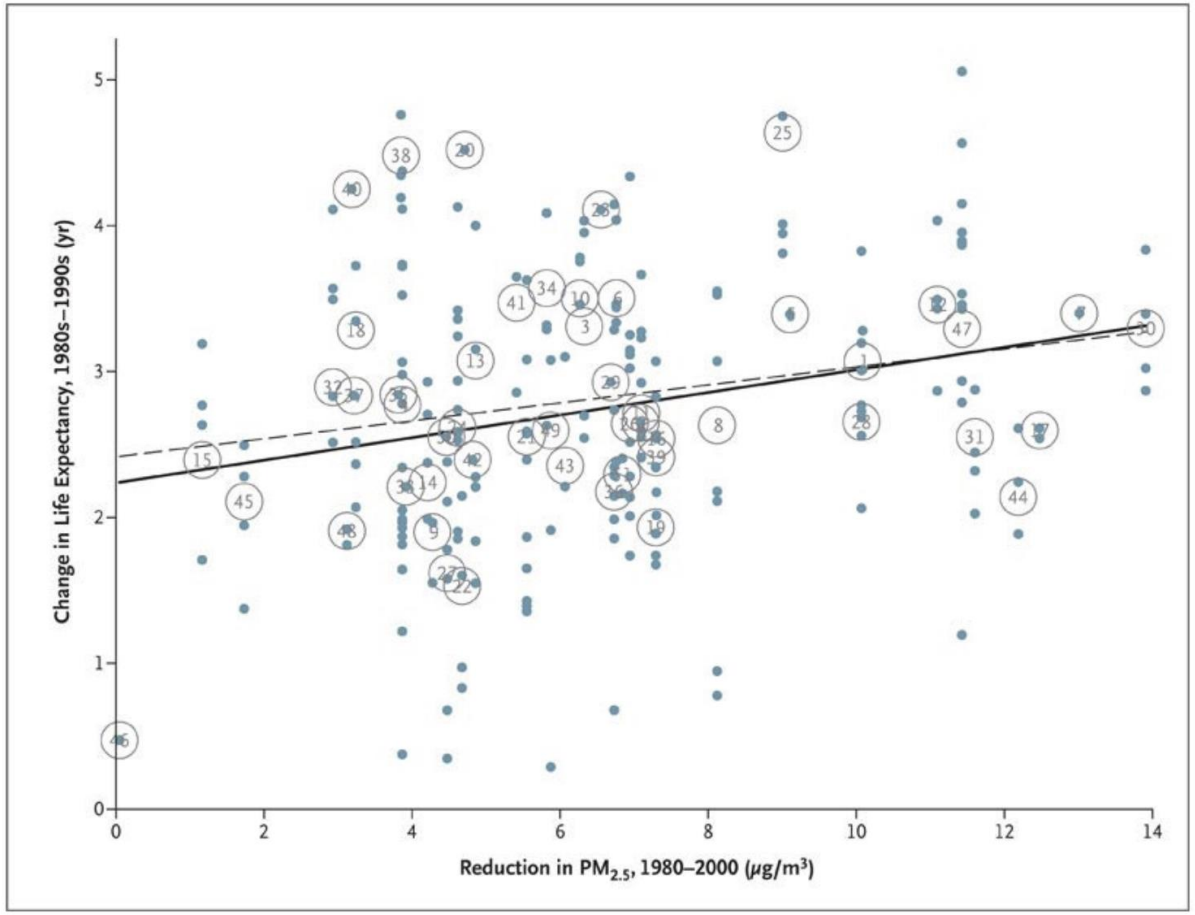
Designs for studying long-term health effects of air pollution

- Main study designs used:
  - Ecological studies
  - Cross-sectional studies
  - Case-control studies
  - Cohort studies
- Early evidence of health impacts of long-term exposures from ecological and cross-sectional studies
- Case-control and cohort studies can be better for causal inference because they capture temporality between exposure and the health outcome

Ecological studies

- Unit of observation is the population or a community/group
- Health outcomes and exposures are measured in the population or community and their association evaluated
- Exposure estimates are a proxy based on the average in the population

Changes in Life Expectancy for the 1980s–1990s, Plotted against Reductions in $PM_{2.5}$ Concentrations for 1980–2000 in the United States counties and metro areas.



See example: Pope III, C. Arden, Majid Ezzati, and Douglas W. Dockery. "Fine-particulate air pollution and life expectancy in the United States." *New England Journal of Medicine* 360, no. 4 (2009): 376-386.

Cohort studies

- Cohort is a group of people with defined characteristics who are followed over time to determine incidence of, or mortality from, a disease, all causes of death, or some other outcome

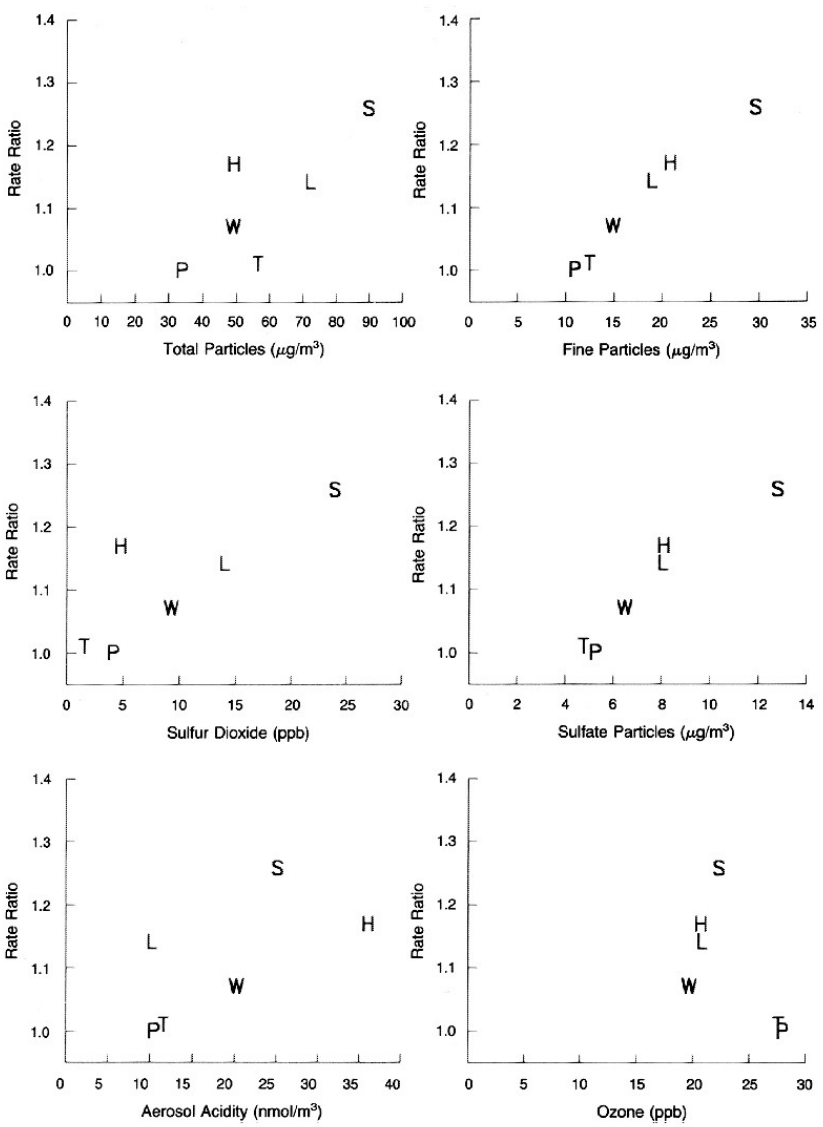
- Key advantage is that it measures the health outcome after exposure, offering a temporal dimension important for causal inference

- Key disadvantage is that you need a large sample size and a potentially long follow-up period → high costs

Harvard Six Cities Study

- Prospective cohort design that followed 8111 adults in 6 U.S. cities for up to 14 years starting in 1974
- Cities were located across the U.S.
- Participants were white and 25-74y at enrolment. Collected data on health and many potential confounders
- Air pollution monitored at central locations

Estimated Adjusted Mortality-Rate Ratios and Pollution Levels in the Six Cities

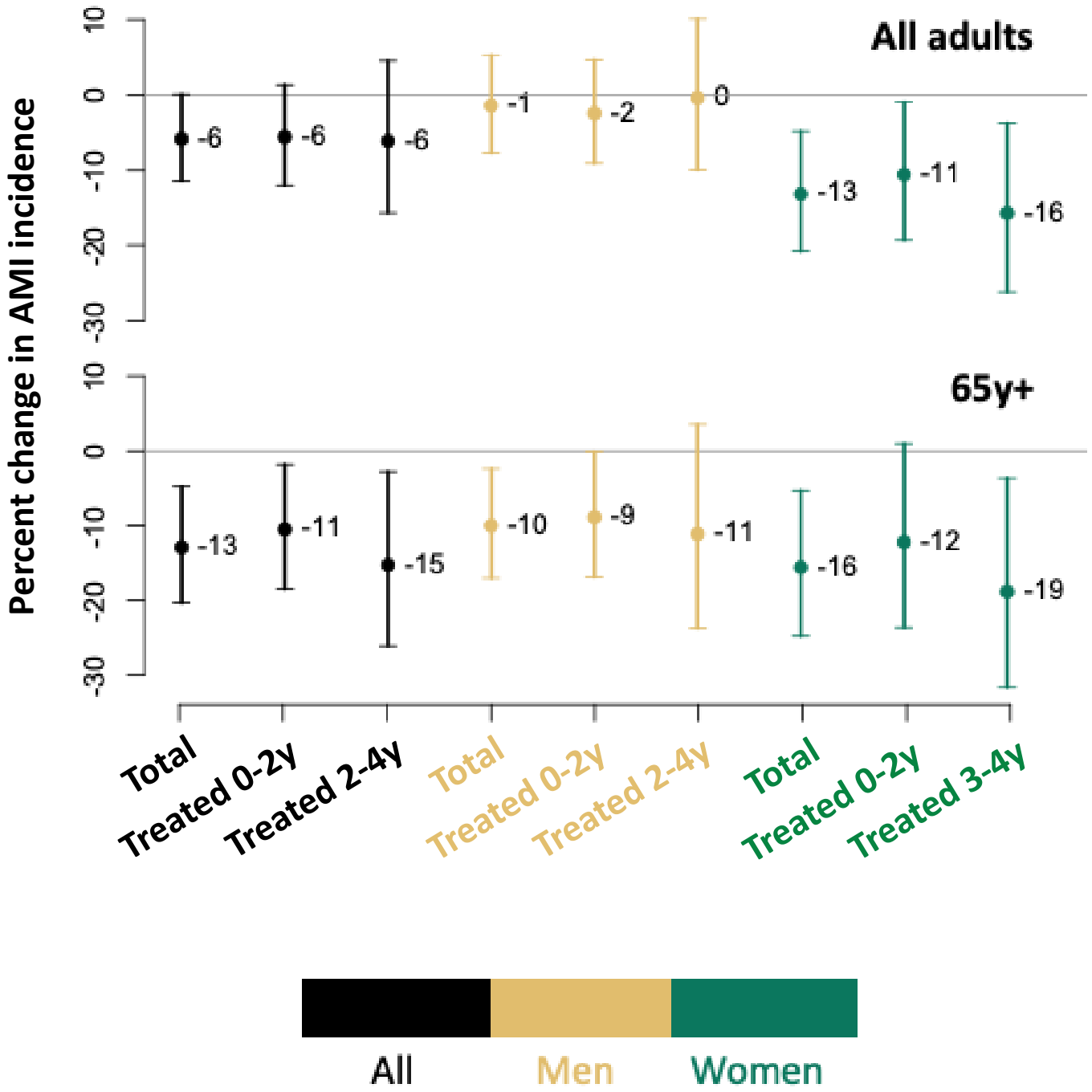


For more reading: Dockery, D. et al, 1993. An association between air pollution and mortality in six US cities. *NEJM*, *329*(24), pp.1753-1759.

Experimental and quasi-experimental designs

- Experimental designs - exposure randomly assigned by the researcher
  - Chamber studies in healthy participants
  - Randomized trials of air pollution interventions (e.g., stoves, air filters)

- Quasi-experimental design – exposure not assigned by the researcher
  - Measures health outcomes before and after a policy or other event that (potentially) changes exposure to air pollution (i.e., powerplant closure, smoking ban), ideally with a control group not exposed to the event.



Group-time average treatment (ATT) effects of the coal-to-clean energy policy on acute myocardial infarction (AMI) in Beijing adults.

Results from adjusted staggered difference-in-difference model

Lee and Chang et al., *under review*

Public health and policy impacts

- Air pollution epidemiology provides new insights that are useful for decision markers and support environmental policy decisions
  - e.g., Observational studies had major influence on US EPA developed National Air Quality Standards (NAAQS) for sex pollutants, China's air quality standards, and the World Health Organization guidelines
  - e.g., Concentration-response studies form the basis for quantitative risk assessment *(to be discussed later in the workshop)*

- Studies from low air pollution countries like Canada demonstrate that health impacts exist below some of the current standards and guidelines, especially interim ones

5 things I wish I knew as an early career researcher*

***I'm still learning some of these myself**

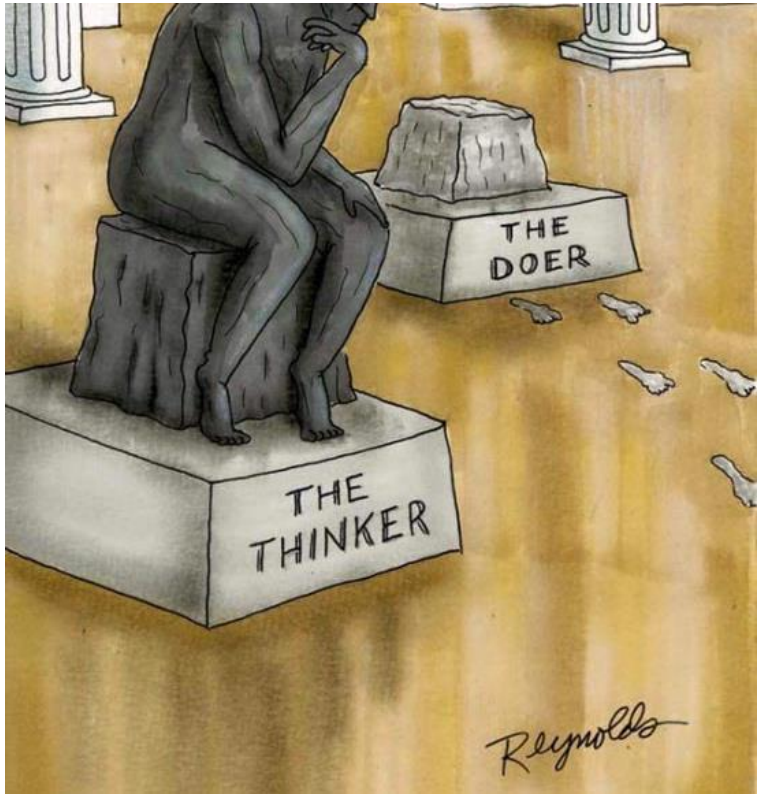
1. **Play to your strengths, work on weaknesses**

1. Play to your strengths, work on weaknesses, an example:

The Thinker

Reads all the relevant literature. Debates and puzzles and works over every single concept and idea in great detail and at great length.

Risk: late or never finished projects; too few outputs, even if those few outputs are great.



The Doer

Sees the task and gets it done – quickly and efficiently.

Risk: may be deemed superficial or feared to be sloppy; can raise suspicion of scientific rigor

1. Play to your strengths, work on weaknesses

Step 1: know where you stand

Step 2: work on your weaknesses

- *Thinkers:* force yourself in the doer direction. Set goals and deadlines and force yourself to meet them. Stop thinking so much, more doing. You will be more productive and publish more papers.
- *Doers*: force yourself in the thinker direction Slow yourself down and do more reading and thinking. Check your work. You will publish better papers.

2. Avoid “method in search of a question’ (for the most part – there are important exceptions)

Publishing good work depends on **good questions** and **good data**. Good data require **good study design** – first focus on those.

- Nothing can fix a poorly designed study.
- Nothing can fix Type 3 error: precise estimates of pointless questions.

Once you have good data to answer a good question, then focus on the fancy modeling.

3. Apply for funding early and often

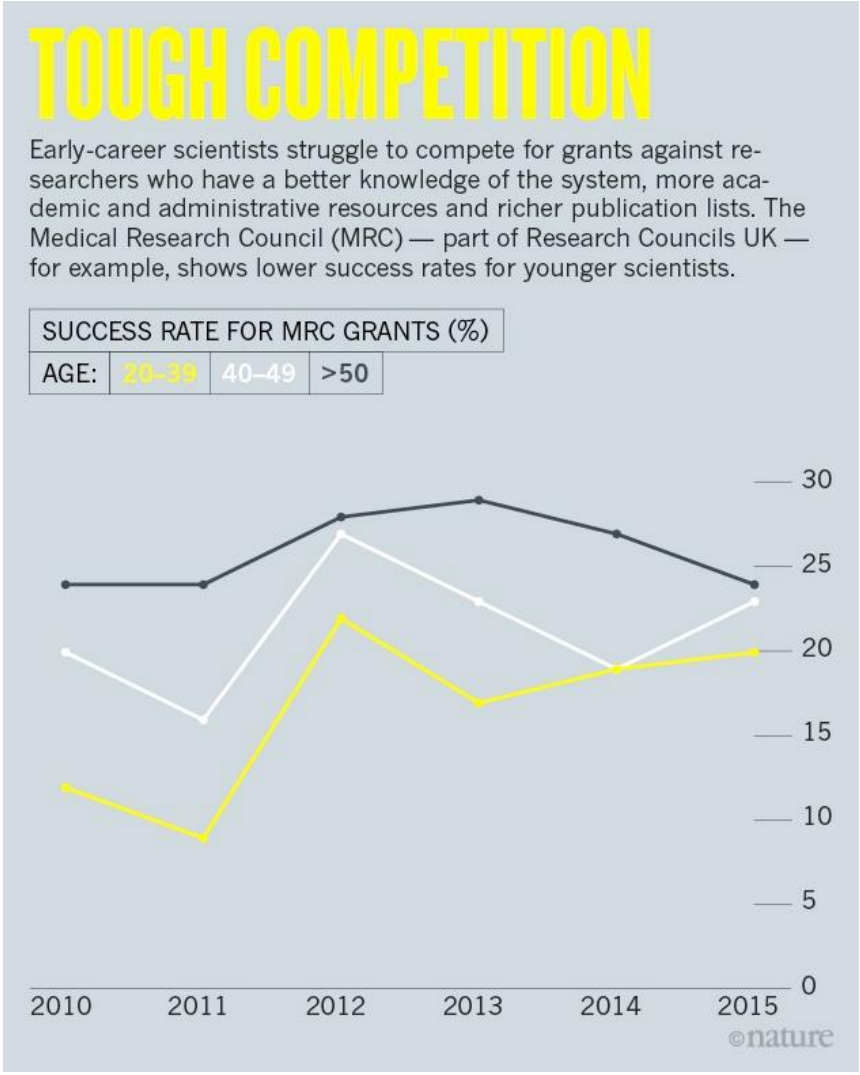
The more grants you write, the higher your probability of success

Rule of increasing returns (in grants)

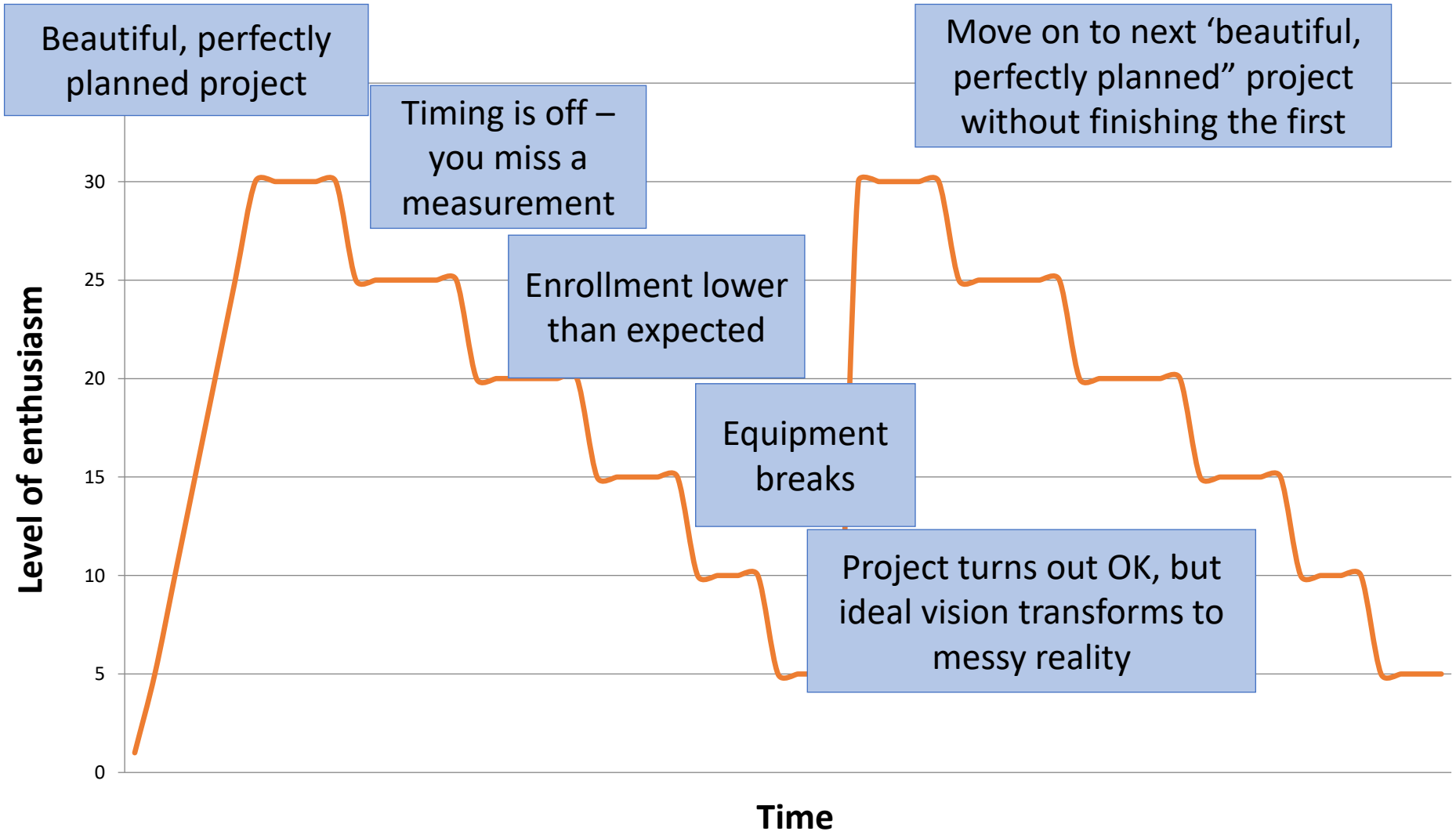
- A $10,000 pilot grant can put you in the running for a $1 million project grant

There is often large scientific value in small pots of money

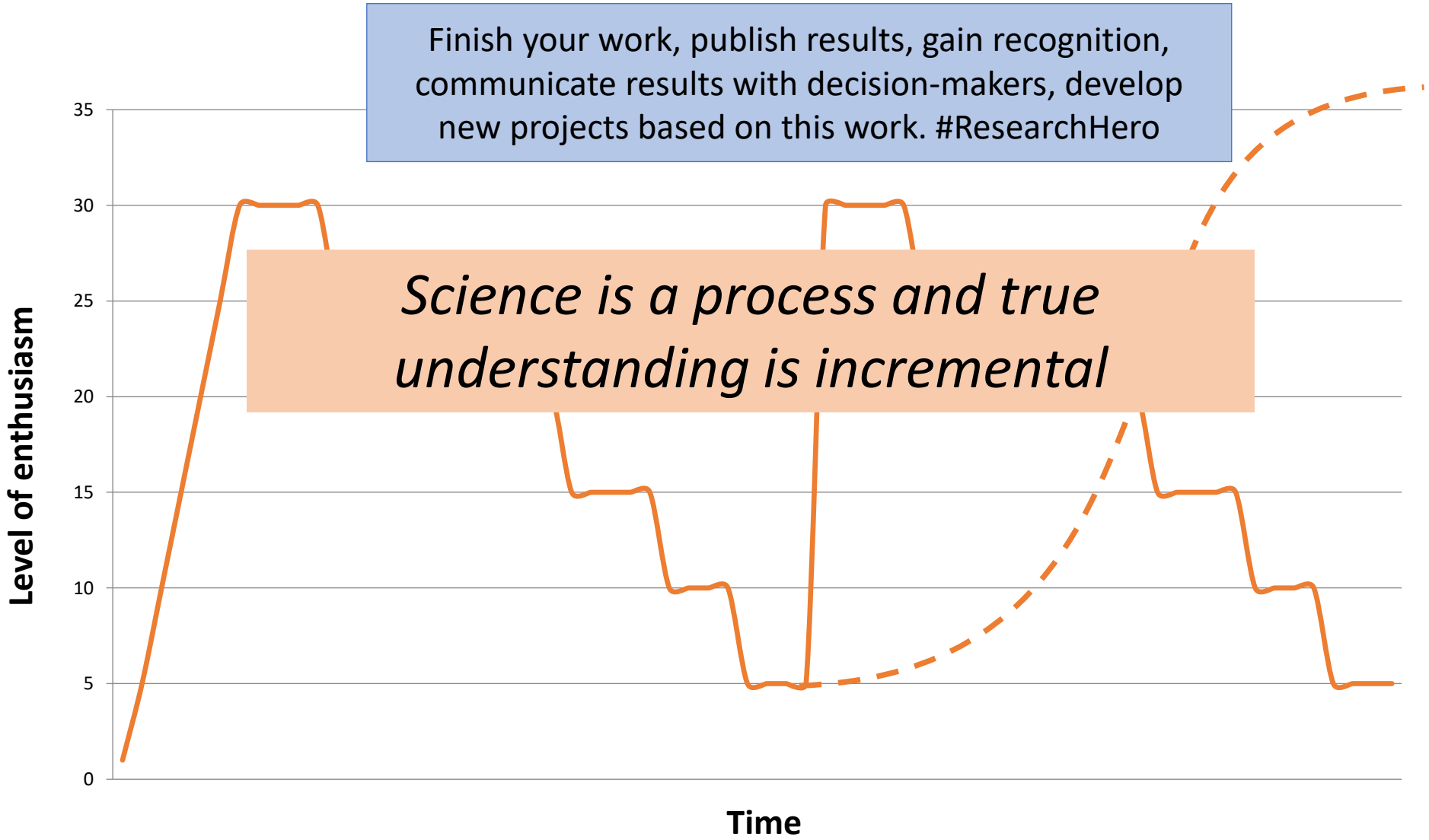
When you receive funding, add it on your CV – even if the actual $ amount is small.



4. Avoid the ‘grass is greener’ syndrome

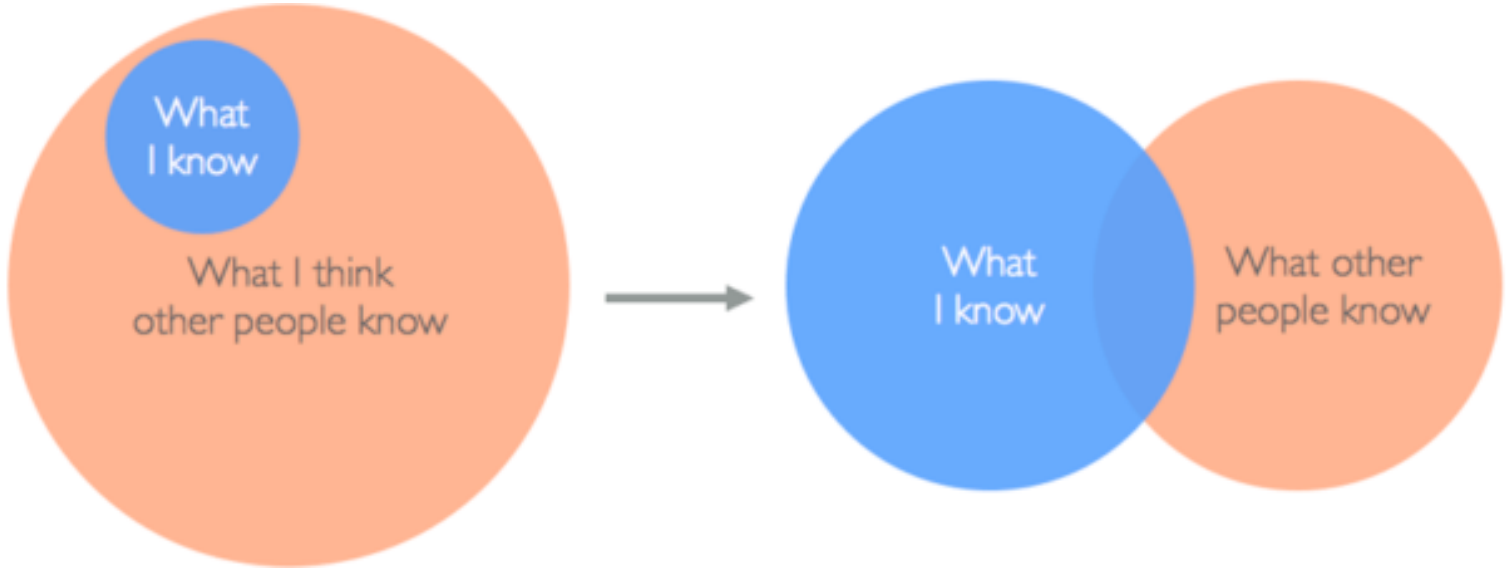


4. Avoid the ‘grass is greener’ syndrome



5. Imposter syndrome is real – move beyond it

Those with imposter syndrome remain convinced that they are frauds and do not deserve their success. Proof of success is dismissed as luck, timing, or a result of deceiving others into thinking they are smarter and more competent than they believe themselves to be. (via WIKIPEDIA)



5. Imposter syndrome is real – move beyond it

Get a fancy fellowship?

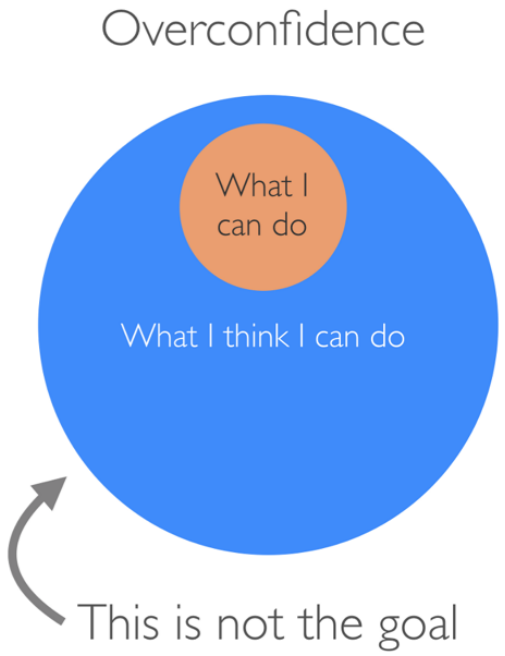
- It's because you deserved it.

Exciting research result?

- It's because you obtained it.

Give a great presentation?

- It's because you practiced it.



Helps to have a supportive university, department, supervisors (if relevant), and group of peers. *But research success requires you.*

Does not preclude you from seeking help from others – nobody is an expert in everything and collaboration is one of most rewarding aspects of research