SCRIPPS INSTITUTION OF OCEANOGRAPHY

QUASI-EXPERIMENTAL METHODS FOR CLIMATE EPIDEMIOLOGY PART I

07 March 2025

Tarik Benmarhnia & Yiqun Ma

University of California, San Diego

Drexel Climate Change and Urban Health Research Center Workshop

OUTLINE

- A general introduction to causal inference
- Study designs for acute health impacts of environmental exposures
  - Time series modelling and case crossover designs
- 2-way Fixed Effect (TWFE) Analyses for air pollution epidemiology

WHAT DO EPIDEMIOLOGISTS DO?

- **Describe** patterns of injuries, disease, and other health outcomes
- **Understand the causes** that produce those patterns [etiology]
  - **Evaluate** the effectiveness of interventions [etiology too]
- **Suggest** [clinical, individual, or population] interventions to solve undesirable patterns

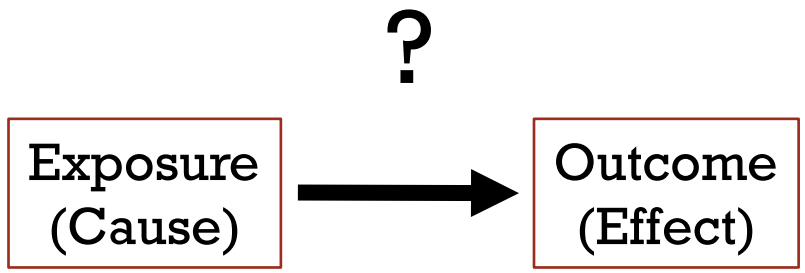
Example:
- Birth weight and PM2.5

A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks

Miguel A. Hernán, John Hsu, and Brian Healy

	Data Science Task		
	Description	**Prediction**	**Causal inference**
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?
Data	• Eligibility criteria • Features (symptoms, clinical parameters …)	• Eligibility criteria • Output (diagnosis of stroke over the next year) • Inputs (age, blood pressure, history of stroke, diabetes at baseline)	• Eligibility criteria • Outcome (diagnosis of stroke over the next year) • Treatment (initiation of statins at baseline) • Confounders • Effect modifiers (optional)
Examples of analytics	Cluster analysis …	Regression Decision trees Random forests Support vector machines Neural networks …	Regression Matching Inverse probability weighting G-formula G-estimation Instrumental variable estimation …

THE CENTRAL NOTION OF CAUSATION



- Different approaches to understand causation in epidemiology
- But the ultimate goal is to estimate the causal effect of a well-defined exposure
  - **Ideally** : an exposure that corresponds directly to a potential intervention

ASSUMING AND ASKING GOOD CAUSAL QUESTIONS

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Miguel A. Hernán, MD, DrPH

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results.

(*Am J Public Health*. 2018;108: 616–619. doi:10.2105/AJPH. 2018.304337)

CORRELATION VS. CAUSATION

Spurious correlations

See: http://www.tylervigen.com/spurious-correlations

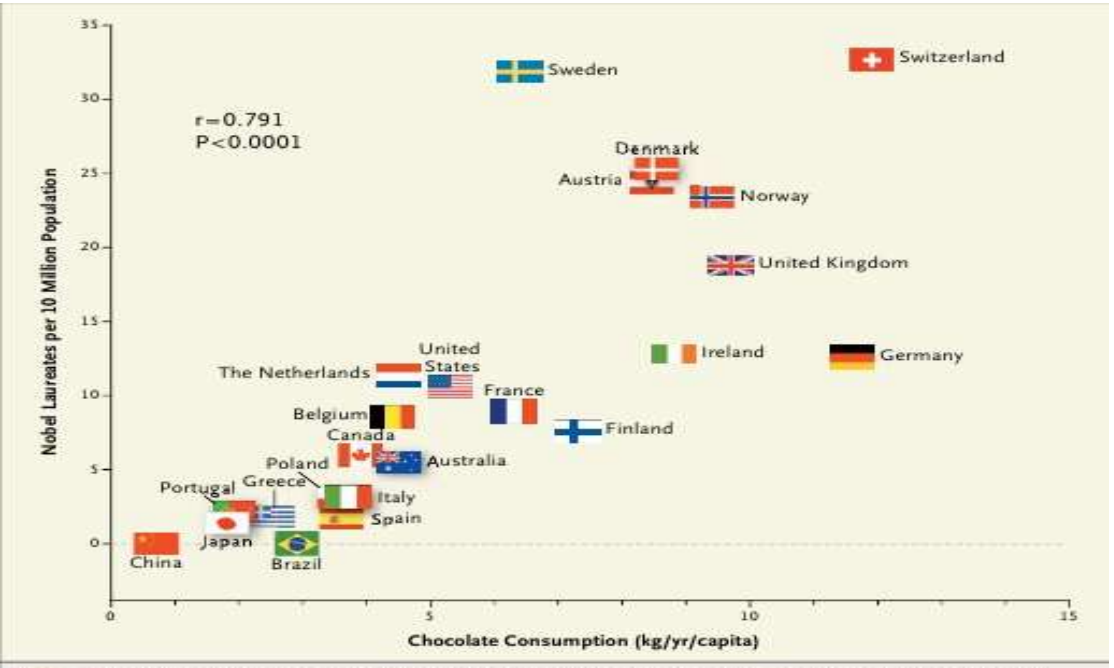
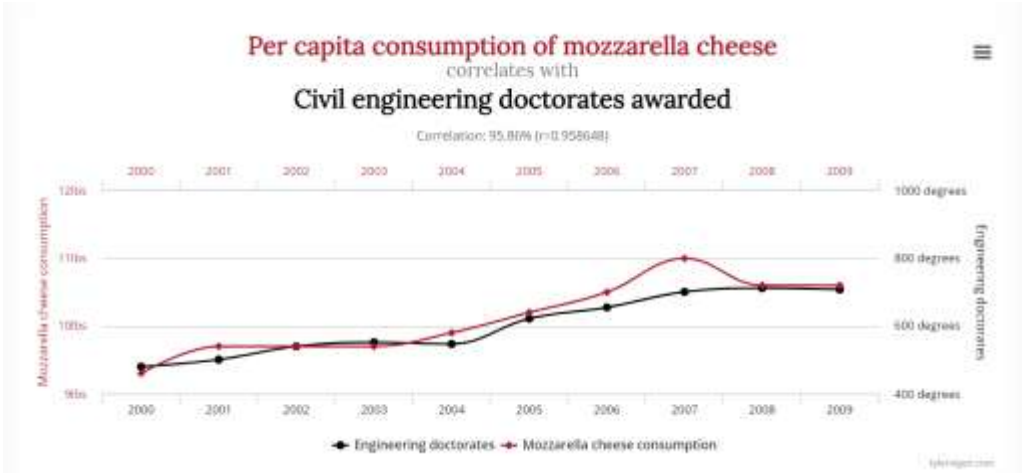
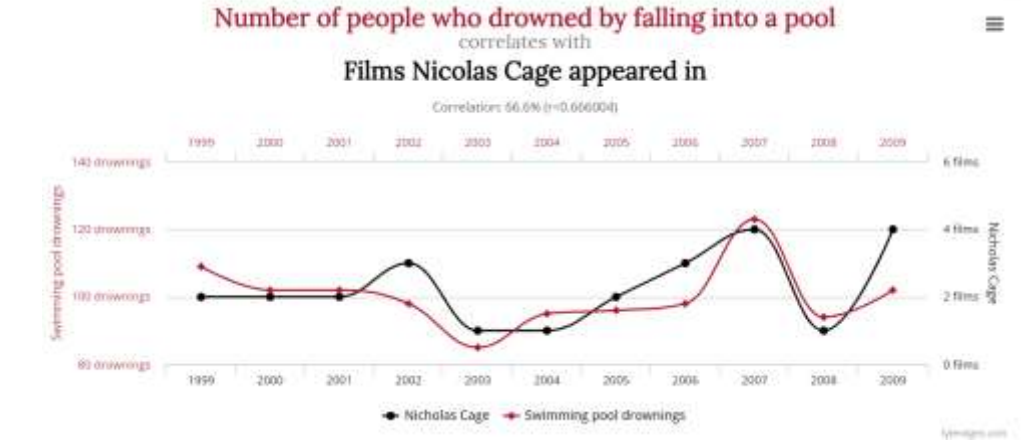


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.





- What can explain the correlation between 2 variables X and Y?
- The importance of conceptualizing a potential intervention (in public health)

"CORRELATION DOES NOT GUARANTEE CAUSATION." YES, BUT...

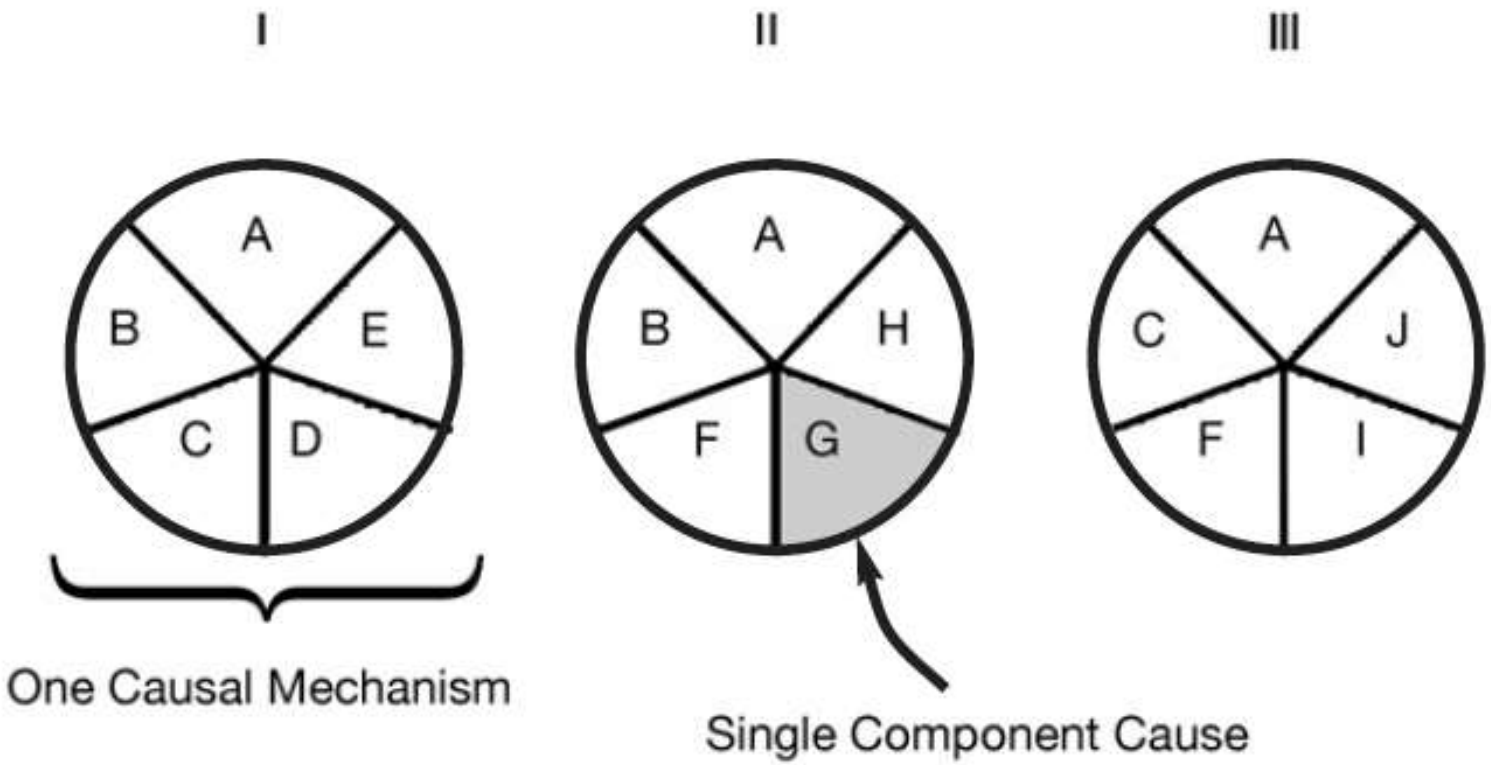
- Correlation is a mathematical concept, causation is not
- There are many reasons to explain an observed correlation/association between 2 variables
- One of them is the causal relationship of interest: X causes Y
- **Other explanations ?**
- The importance of the word "inference" and a causal framework

HILL'S CAUSAL CRITERIA (1965)

1. Strength
2. Consistency
3. Specificity
4. Temporality
5. Biologic gradient
6. Plausibility
7. Coherence
8. Experimental evidence
9. Analogy

Is it always relevant?

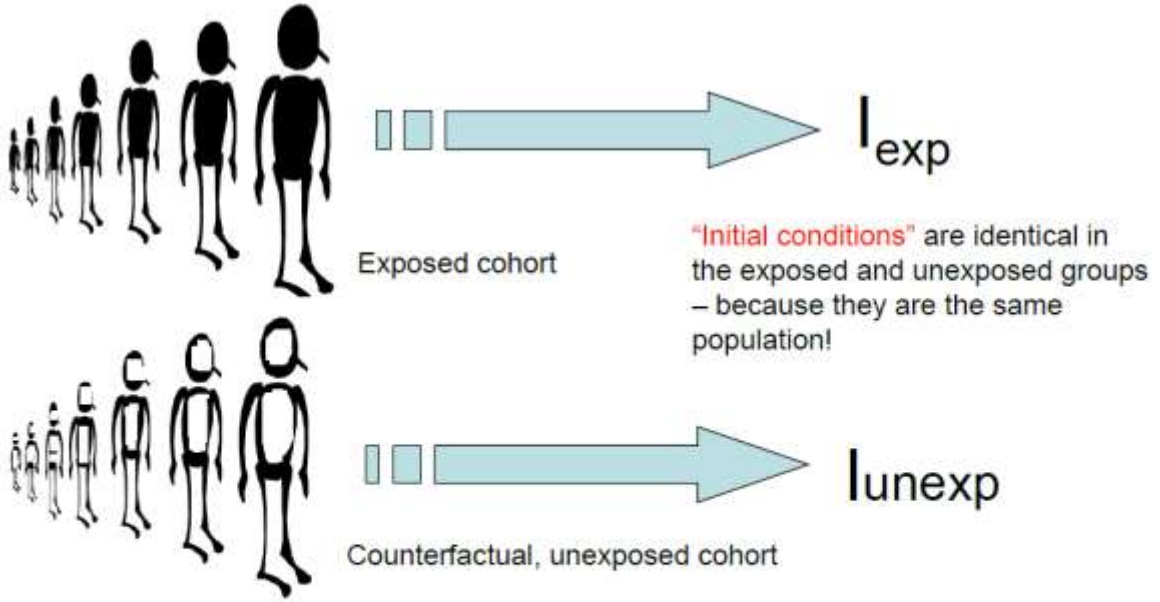
ROTHMAN'S SUFFICIENT-COMPONENT CAUSES [CAUSAL PIES] (1976)



THE COUNTERFACTUAL MODEL

- One of the most common model used to identify causal effects is based on the notion of hypothetical counterfactuals
- Within the counterfactual model, causal effects are defined as a contrast of two outcome values, one observed (factual) and one unobserved (counterfactual)
- The fundamental problem of causal inference is that, by definition, the counterfactual outcome is impossible to observe (Holland, 1986)

IDEAL CAUSAL CONTRAST

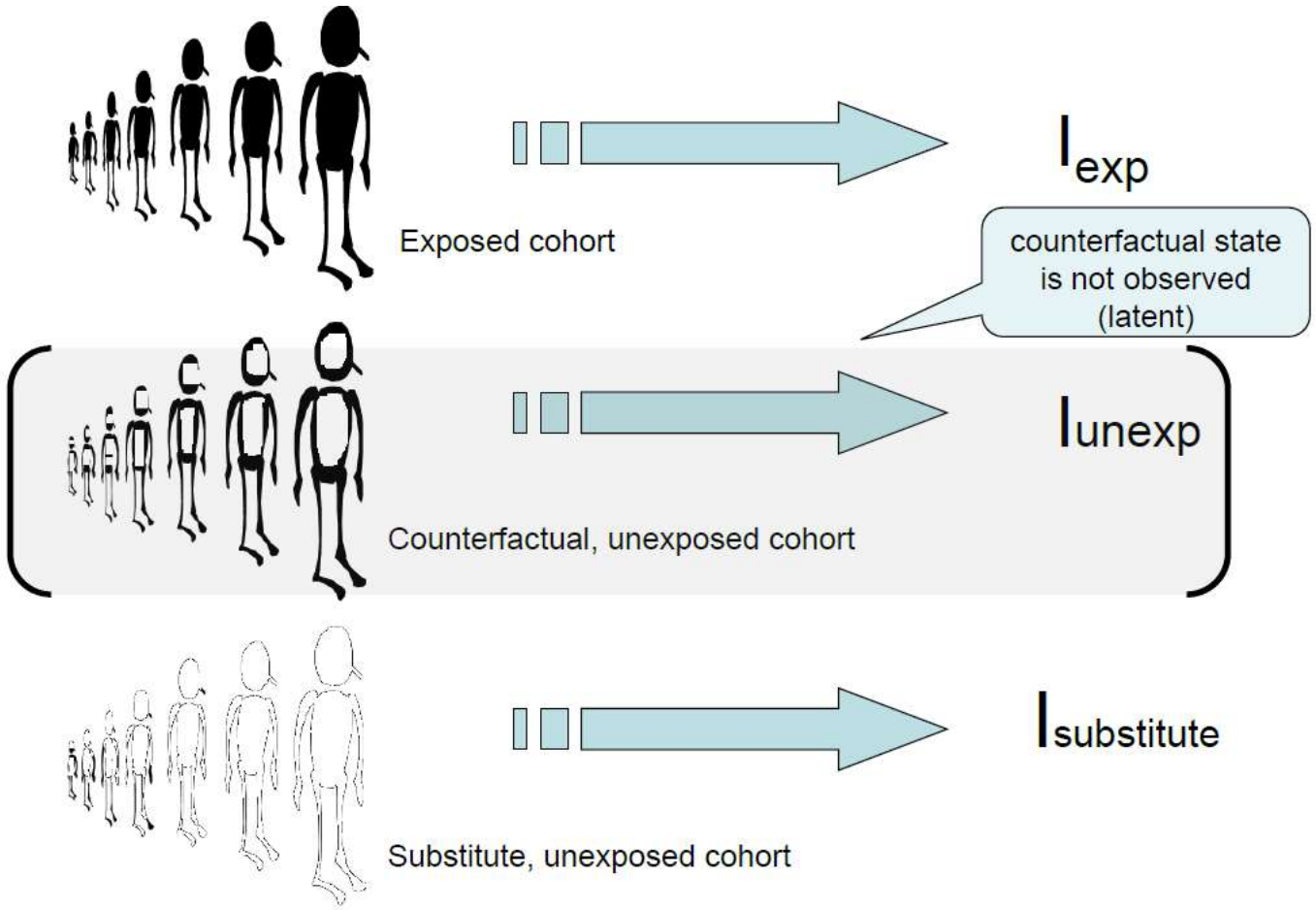


"A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period"

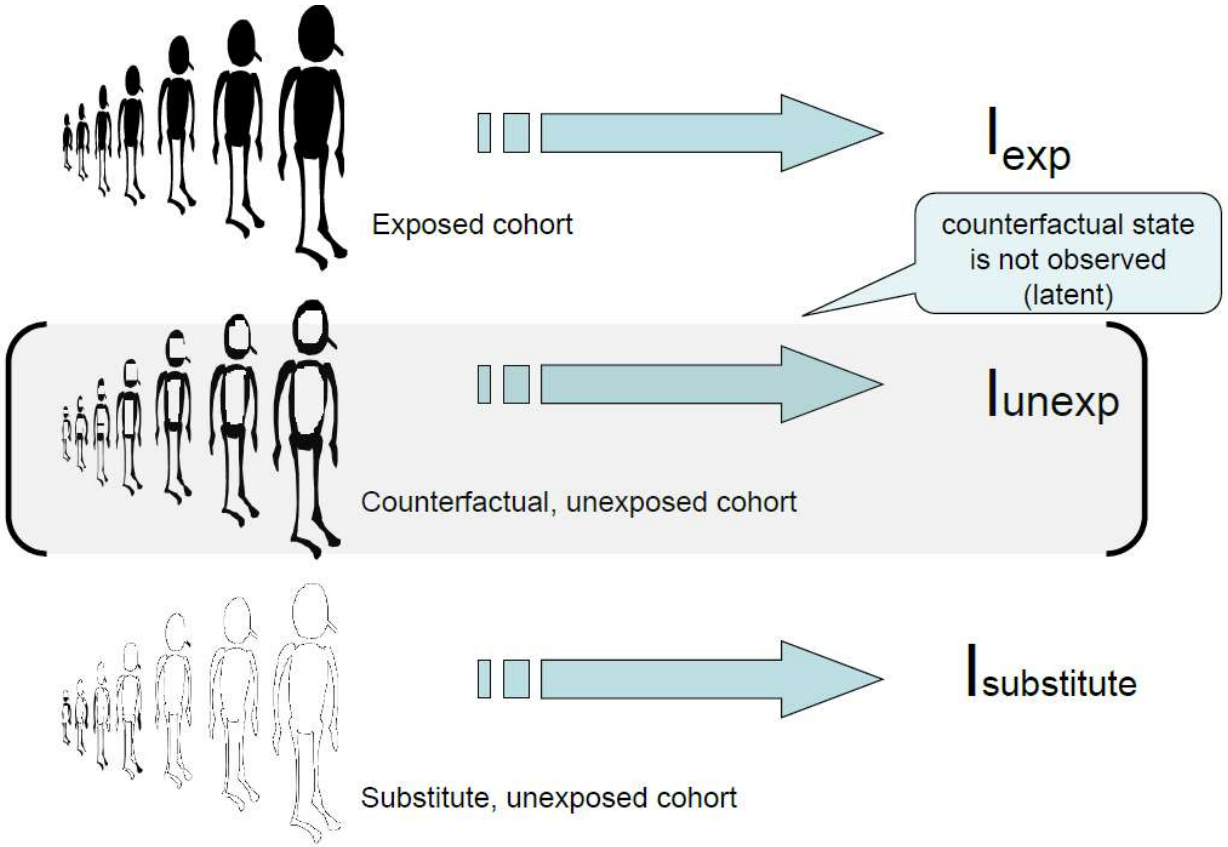
Maldonado & Greenland, Int J Epi 2002;31:422-29

The **fundamental problem** of causal inference is that, by definition, the counterfactual outcome is impossible to observe.

COUNTERFACTUALS IN PRACTICE



FINDING A SUBSTITUTE FOR THE COUNTERFACTUAL GROUP



Some Notations

$$\textbf{(i)}\quad \Pr[Y^{a=1}=1]-\Pr[Y^{a=0}=1]=0$$

$$\textbf{(ii)}\quad \frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}=1$$

$$\textbf{(iii)}\quad \frac{\Pr[Y^{a=1}=1]/\Pr[Y^{a=1}=0]}{\Pr[Y^{a=0}=1]/\Pr[Y^{a=0}=0]}=1$$

MAIN ASSUMPTIONS IN THE COUNTERFACTUAL FRAMEWORK

- **Exchangeability**
- **Positivity**
- **Consistency**
  - **Stable unit treatment value assumption (SUTVA)**

EXCHANGEABILITY

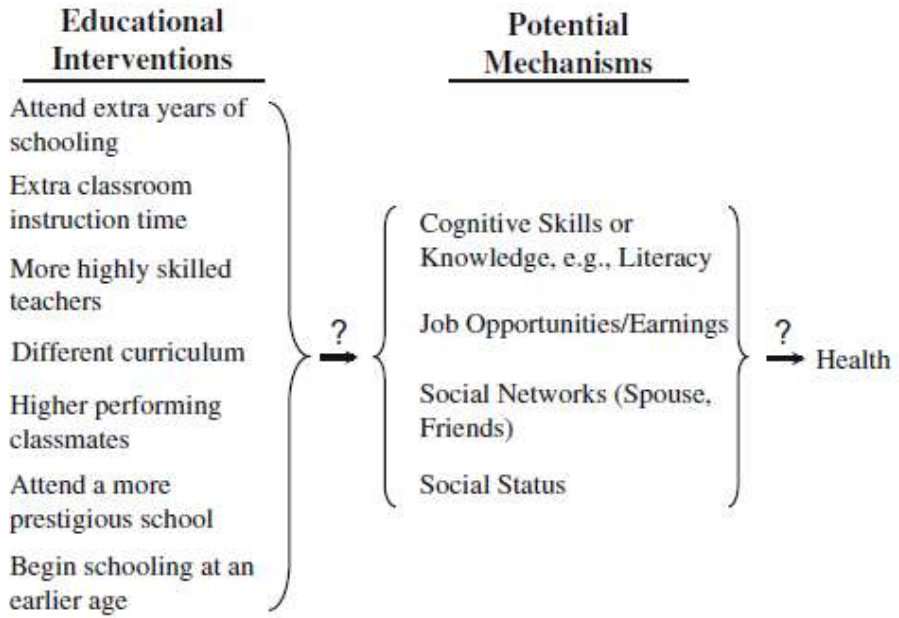
- In the presence of exchangeability, the counterfactual risk under treatment in the white part of the population would equal the counterfactual risk under treatment in the entire population
- It means that the counterfactual outcome and the actual treatment are independent
- In other words, we could **substitute** exposed and unexposed groups and would get the same targeted effect
  - P. 27: "*the treated and the untreated are exchangeable because the treated, had they remained untreated, would have experienced the same average outcome as the untreated did, and vice versa*".
- We thus need to reach **covariate balance**
- For both measured and unmeasured confounders

POSITIVITY

- Individuals in every stratum of the covariates have a non-zero probability of being in the exposed group and of being in the unexposed group
- Sometimes referred to as the experimental treatment assumption
- Consider a basic example with a binary treatment A
  - Pr [A = 1] and Pr [A = 0] need to be greater than 0
  - Also Pr [A =a | L = l] > 0 for for all l with Pr [L = l] ≠ 0
  - Otherwise, it will be impossible to assess the effect of A on a given outcome Y
- It may also lead to exchangeability issues
- Examples:
  - Social epidemiology
  - Environmental Epidemiology

CONSISTENCY

- The presence of multiple versions of treatment is problematic:
  1. If the causal effect varies across versions: the magnitude of the average causal effect (ACE) depends on the proportion of individuals who received each version
  2. For the interpretation of the causal effect of interest
- Especially in observational studies
  - Investigators have no control over the versions of the treatment
- Can be solved by restriction (see p.33).
  - Problem 1: sample size
  - Problem 2: Sometimes, there is no data
- (Re) read Hernán and Taubman (2008) on Obesity
- When a Rose is not a Rose: the example of education
- **Well-defined interventions are a pre-requisite for causal inference**



(Rehkopf, Glymour and Osypuk, 2016)

SUTVA

- Stable unit treatment value assumption (SUTVA) implies
  - i) no interference between units
  - ii) access to a unique treatment
- Examples ?

REVISITING STUDY DESIGNS IN EPIDEMIOLOGY

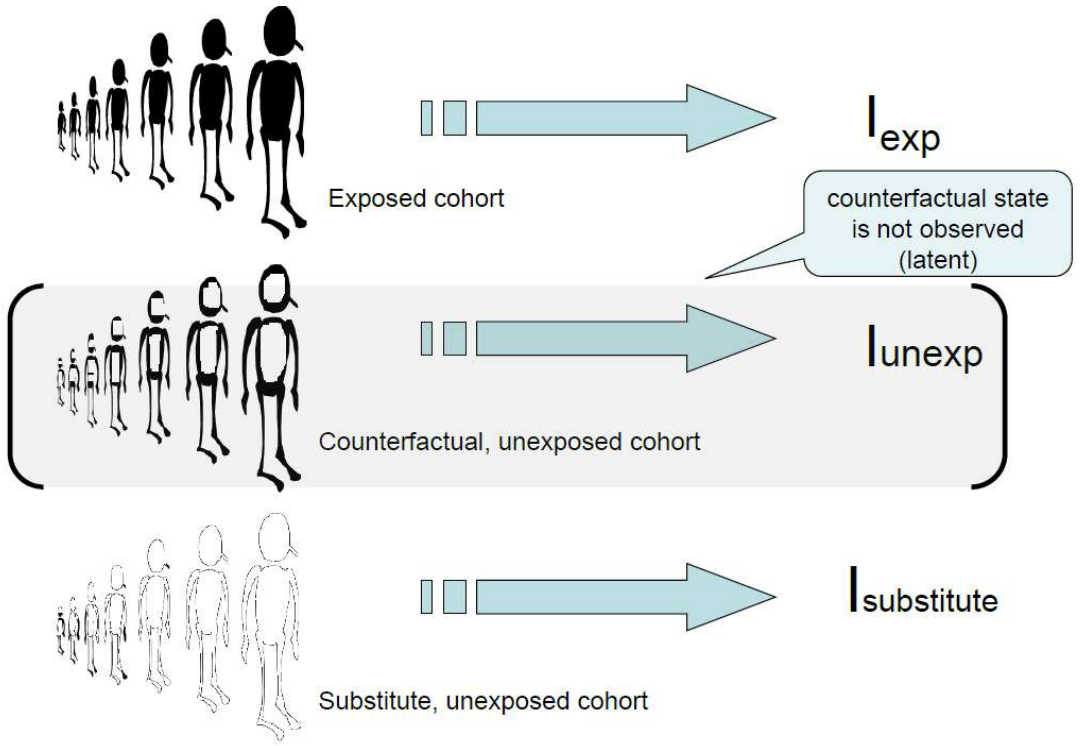
- Experiments: Randomized Controlled Trials
- Observational Cohort studies
- Observational Case control studies
- Observational cross sectional studies

A. At the Individual level
B. At the ecological level
C. Both: Multilevel

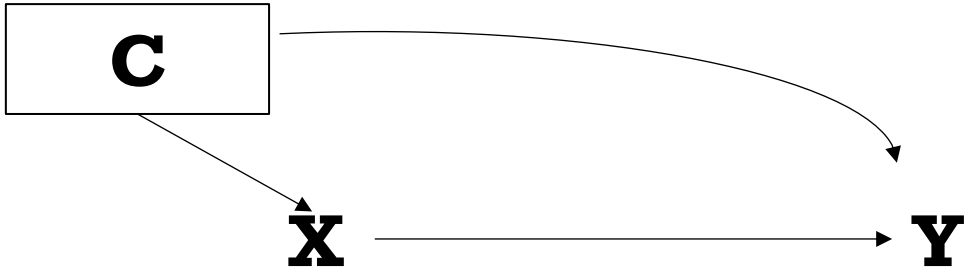
Example: PM2.5 and asthma

BIAS IN EPIDEMIOLOGY

- General definition:
  - a **<u>systematic</u>** error in the design or conduct of a study, resulting from flaws either in the method of selection of study participants or in the procedures for gathering and considering relevant exposure and/or disease information
  - **<u>Internal validity</u>**
- 3 main types of biases
  - Information/misclassification Bias
  - **Selection Bias**
  - **Confounding Bias**



CONFOUNDING BIAS



- Some definitions:
  - Confounding is a type of bias that arises from specific causal structures, typically a shared cause of X and Y
  - If **conditioning** upon a set of covariates Z will render the association between X and Y unconfounded, then Z is a sufficient set of covariates for estimating the X-Y relationship
  - Back-door path to be closed
- **Conditioning** on a covariate:
  - Restriction
  - Sample selection (can be unknown or involuntary)
  - Matching
  - Stratification
  - Regression Adjustment
- Standardization, g-formula
- IP weighting

SELECTION BIAS

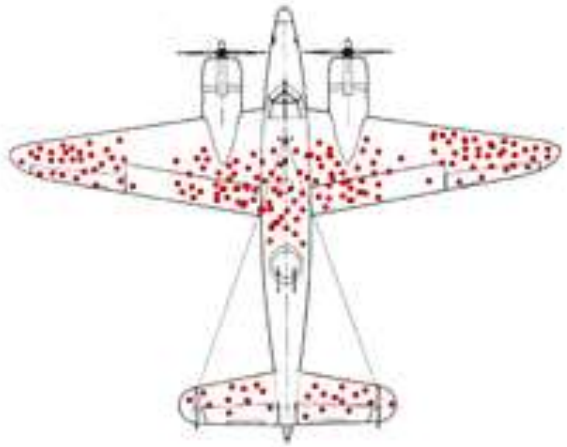
The example of Basketball players

X is height

Y is speed

Z is Basketball skills

- Short people are not necessarily fast, nor are tall people;
- However, if you look only at professional basketball players, you may find that the short ones are very fast
- By **restricting to pro basketball players**, you have conditioned on a common effect of height and speed, and within this stratum of pro ball players, height and speed are (inversely) associated.

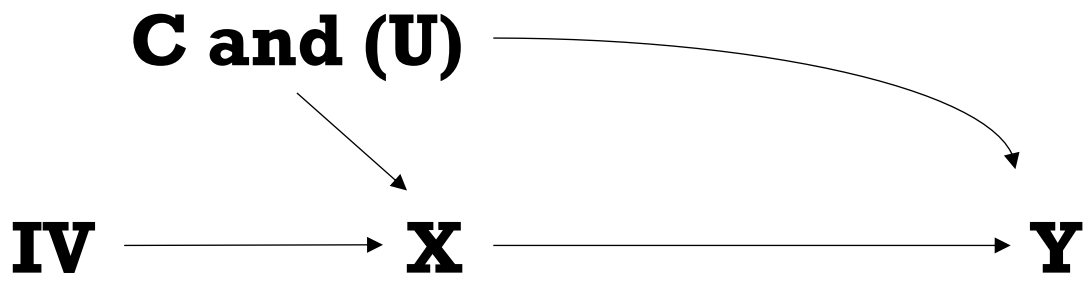


The birthweight paradox

Ellie Murray @EpiEllie 14 juil. 2018

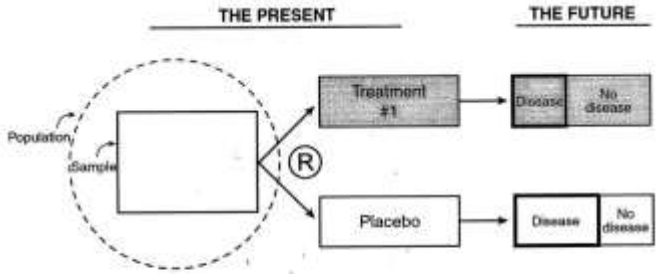
THE IDEA BEHIND RANDOMIZATION

- Randomization can be analyzed as an **Instrumental Variable**
- **So Natural Experiments can be exploited to do the same**
- The overall aim is to deal with measured and unmeasured confounding in observational studies by using an instrumental variable



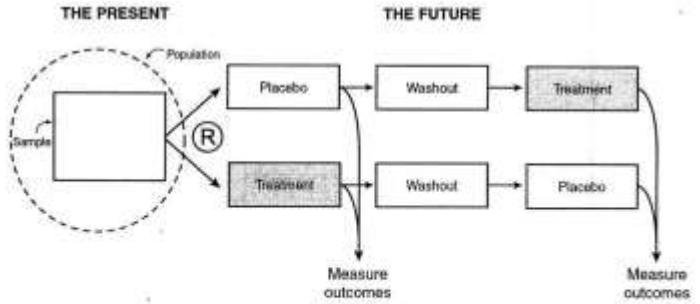
DIFFERENT TYPES OF RCTS

- Two-arm, parallel design
- Planned cross-over design
- Factorial design
- Cluster randomized trials
- **Stepped Wedge Designs**



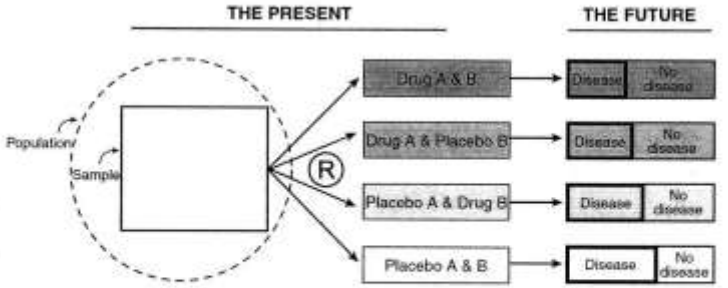
■ **FIGURE 10.1**
In a randomized trial, the investigator (a) selects a sample from the population, (b) measures baseline variables, (c) randomizes the participants, (d) applies interventions (one should be a blinded placebo, if possible), (e) follows up the cohort, (f) measures outcome variables (blindly, if possible) and analyzes the results.

Two-arm, parallel design



■ **FIGURE 11.4**
In the cross-over randomized trial, the investigator (a) selects a sample from the population, (b) measures baseline variables, (c) randomizes the participants, (d) applies interventions, (e) measures outcome variables, (f) allows washout period to reduce carryover effect, (g) applies intervention to former placebo group, (h) measures outcome variables again.

Planned cross-over design



■ **FIGURE 11.2**
In a factorial randomized trial, the investigator (a) selects a sample from the population; (b) measures baseline variables; (c) randomly assigns two active interventions and their controls to four groups, as shown; (d) applies interventions; (e) follows up the cohorts; (f) measures outcome variables.

Factorial design

Hulley et al. Designing Clinical Research. 2nd Edition. Lippincott Williams & Wilkins, 2001

WHY RCT ARE NOT SUFFICIENT?

- The parachute use example
- RCTs are sometimes impossible for:
  - Logistical reasons
  - Ethical reasons
- Sometimes it is too late
- Other misconceptions about RCTs

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Understanding and misunderstanding randomized controlled trials

Angus Deaton[a,b,c,*], Nancy Cartwright[d,e]

BMJ VOLUME 327 20–27 DECEMBER 2003 bmj.com

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

Abstract

Objectives To determine whether parachutes are effective in preventing major trauma related to gravitational challenge.

Design Systematic review of randomised controlled trials.

Data sources: Medline, Web of Science, Embase, and the Cochrane Library databases; appropriate internet sites and citation lists.

Study selection: Studies showing the effects of using a parachute during free fall.

Main outcome measure Death or major trauma, defined as an injury severity score > 15.

Results We were unable to identify any randomised controlled trials of parachute intervention.

Conclusions As with many interventions intended to prevent ill health, the effectiveness of parachutes has not been subjected to rigorous evaluation by using randomised controlled trials. Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.

THE TARGET TRIAL FRAMEWORK

- Target trial emulation is the application of design principles from randomized trials to the analysis of observational data, thereby explicitly tying the analysis to the trial it is emulating.
- The purpose is to improve the quality of observational epidemiology through the application of trial design principles, even when, or perhaps especially when, a comparator trial is not yet available or feasible

Eur J Epidemiol (2017) 32:473–475
DOI 10.1007/s10654-017-0293-4

COMMENTARY

Target trial emulation: teaching epidemiology and beyond

Jeremy A. Labrecque[1] · Sonja A. Swanson[1]

Practice of Epidemiology

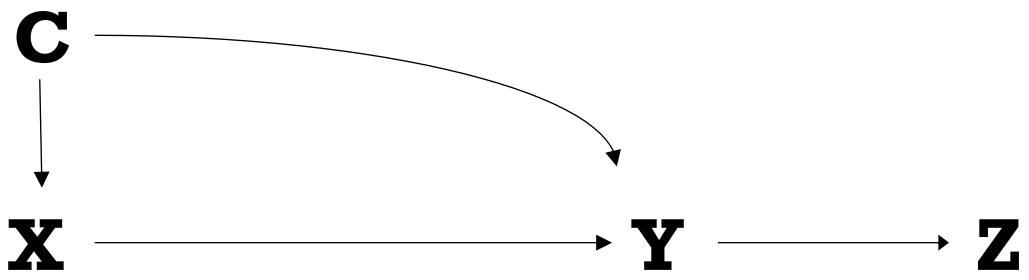
Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

USING DAGS IN EPIDEMIOLOGY

DRAWING A CAUSAL DAG

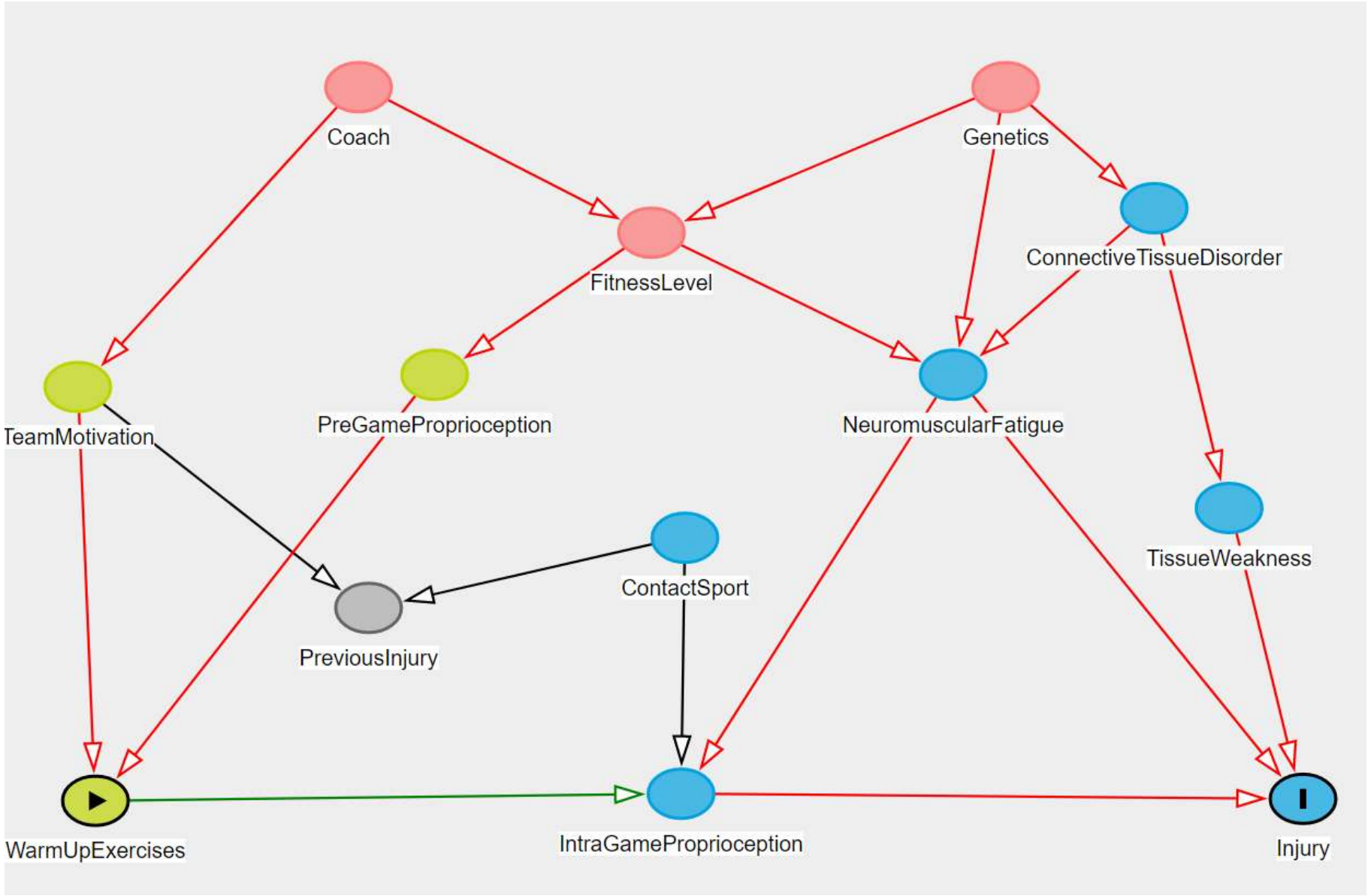
- DAGs are a graphical tool to draw causal assumptions and identify optimal analytical strategies
- Very simple rules



DAGS FACILITATE SEVERAL TASKS

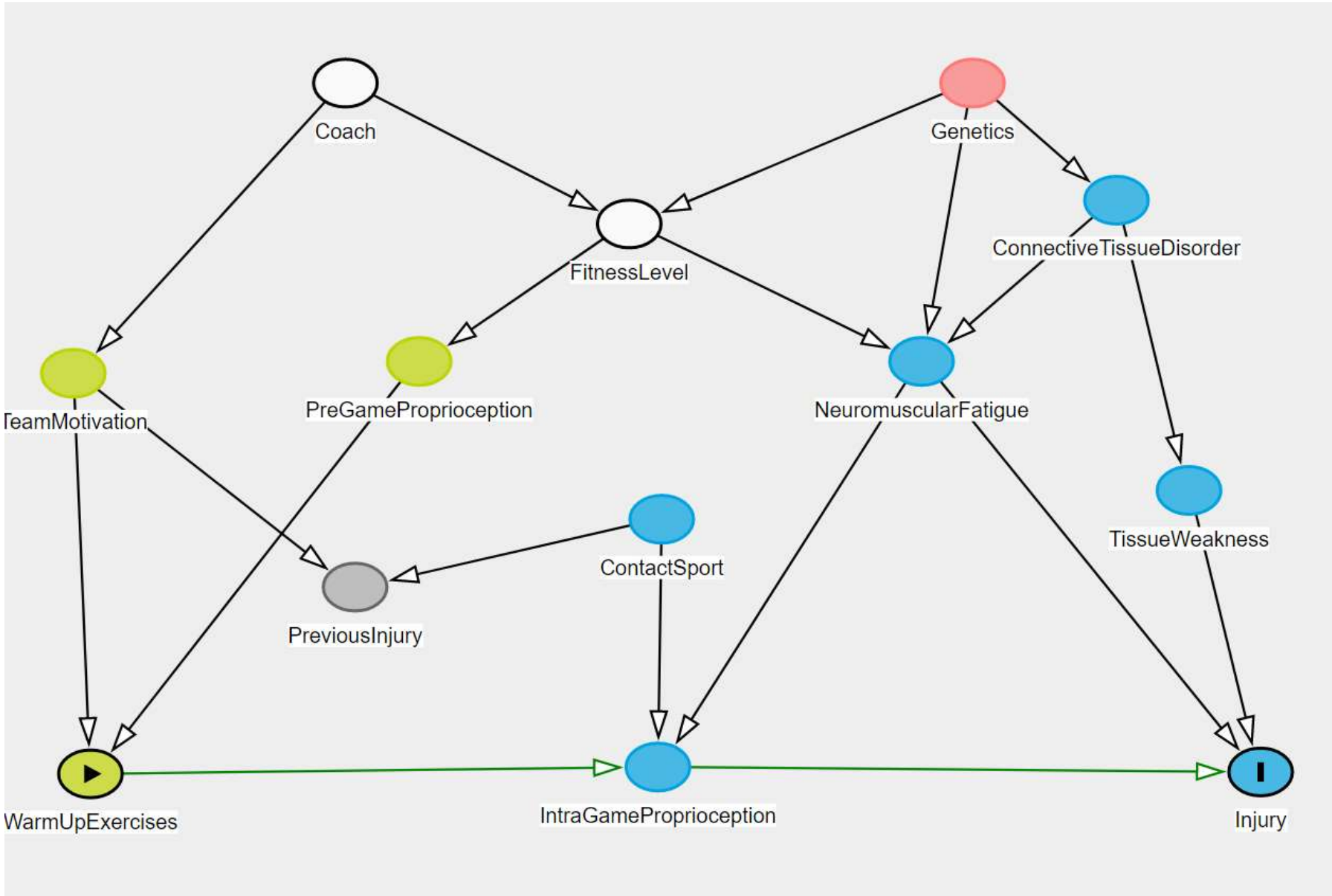
- Choosing covariates to condition on to address confounding
- Optimize such covariate selection
- Recognizing selection bias
- Identify a potential collider bias

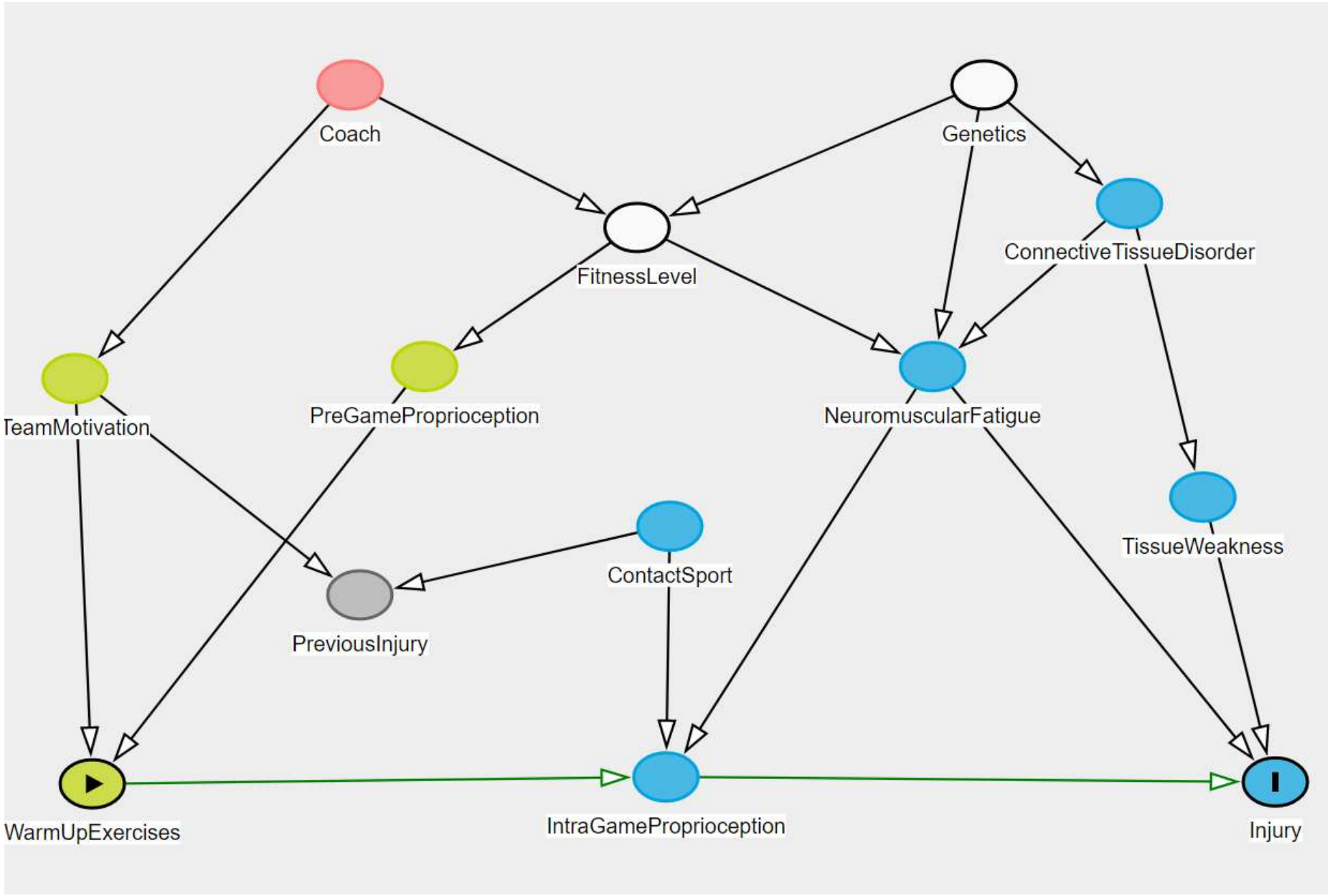
EXAMPLES



Minimal sufficient adjustment sets for estimating the total effect of WarmUpExercises on Injury:

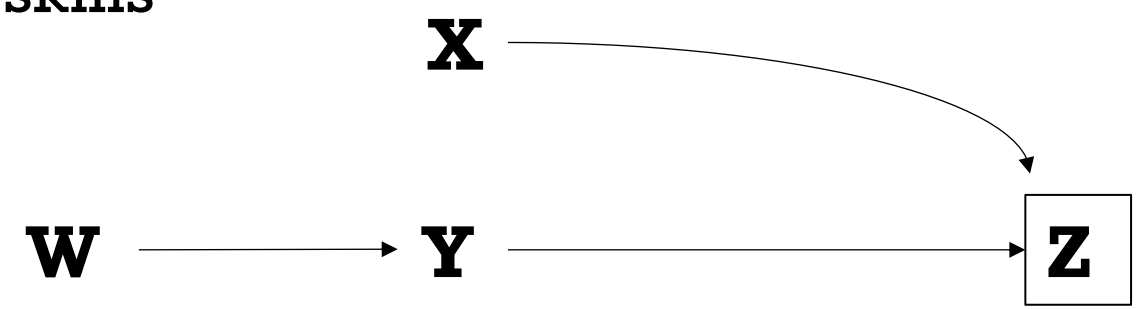
- Coach, FitnessLevel
- Coach, PreGameProprioception
- ConnectiveTissueDisorder, NeuromuscularFatigue
- FitnessLevel, Genetics
- FitnessLevel, TeamMotivation
- NeuromuscularFatigue, TissueWeakness
- PreGameProprioception, TeamMotivation


Coach
Genetics
FitnessLevel
ConnectiveTissueDisorder
TeamMotivation
PreGameProprioception
NeuromuscularFatigue
TissueWeakness
ContactSport
PreviousInjury
WarmUpExercises
IntraGameProprioception
Injury


Coach
Genetics
FitnessLevel
ConnectiveTissueDisorder
TeamMotivation
PreGameProprioception
NeuromuscularFatigue
TissueWeakness
ContactSport
PreviousInjury
WarmUpExercises
IntraGameProprioception
Injury

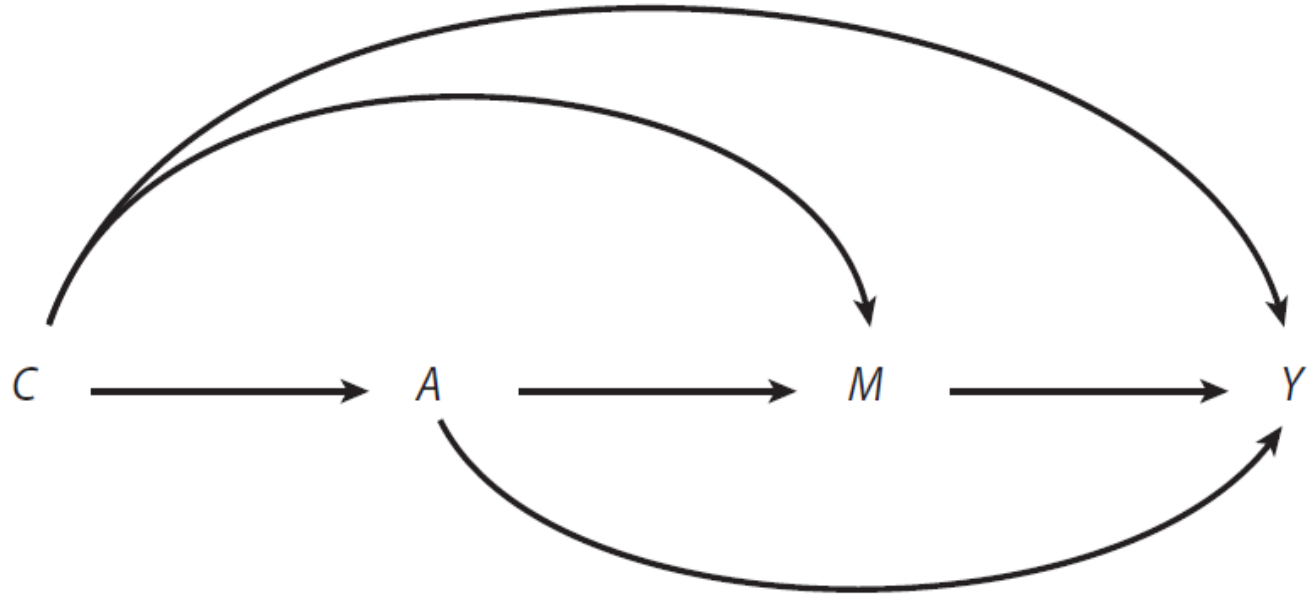
COLLIDER BIAS

- When conditioning on a shared effect of X and Y
  - Selection Bias
  - Incorrect conditioning
- The example of Basketball players
  - X is height
  - Y is speed
  - Z is Basketball skills



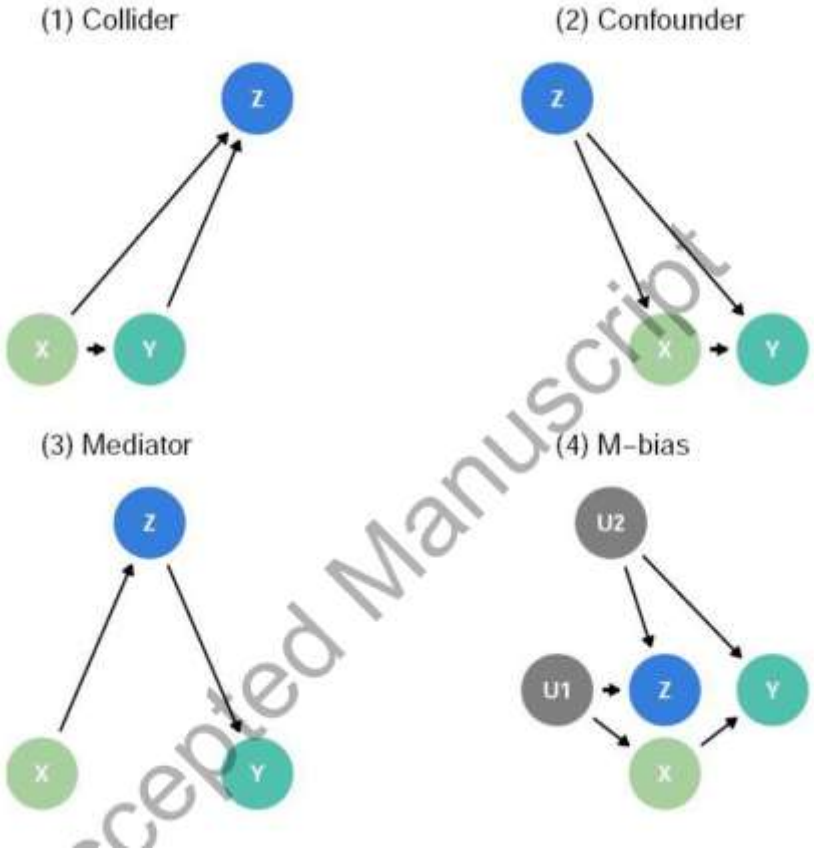
ADJUSTING ON A MEDIATOR

- Mostly a problem of interpretation
  - But can also lead to collider stratification bias
- Controlled Directed Effect instead of Total Effect



Causal inference is not just a statistics problem

Lucy D'Agostino McGowan[a,*], Travis Gerke[b], Malcolm Barrett[c]



Data generating mechanism	Correct causal model	Correct average causal effect
(1) Collider	Y ~ X	1
(2) Confounder	Y ~ X ; Z	0.5
(3) Mediator	Direct effect: Y ~ X ; Z Total Effect: Y ~ X	Direct effect: 0 Total effect: 1
(4) M-Bias	Y ~ X	1

Directed Acyclic Graphs describing the four data generating mechanisms: (1) Collider (2) Confounder (3) Mediator (4) M-Bias.

Lucy D'Agostino McGowan, Travis Gerke & Malcolm Barrett (01 Dec 2023): Causal inference is not just a statistics problem, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2023.2276446

A QUICK OVERVIEW ON EFFECT MEASURE MODIFICATION

EFFECT MODIFICATION

- We say that M is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of M.
- Since the average causal effect can be measured using different effect measures (e.g., risk difference, risk ratio), the presence of effect modification depends on the effect measure being used
  - This is why we talk about effect measure modification

Additive effect modification:

$$\mathrm{E}[Y^{a=1} - Y^{a=0} | M = 1] \neq \mathrm{E}[Y^{a=1} - Y^{a=0} | M = 0]$$

Multiplicative effect modification:

$$\frac{\mathrm{E}[Y^{a=1}|M=1]}{\mathrm{E}[Y^{a=0}|M=1]} \neq \frac{\mathrm{E}[Y^{a=1}|M=0]}{\mathrm{E}[Y^{a=0}|M=0]}$$

THE DIFFERENCE BETWEEN CONFOUNDING AND EFFECT MEASURE MODIFICATION

- We want to condition on or control/adjust for confounding
  - This is a bias
  - Using different techniques
- EMM is not a bias so we do not need to control/adjust for it
- A variable can be both a confounder and an EMM
- Instead, we assess EMM to better understand the mechanisms underlying a specific causal association

Example with RCTs

THE DIFFERENCE BETWEEN INTERACTION AND EFFECT MODIFICATION

- The concept of effect (measure) modification refers to the causal effect of A, not to the causal effect of E
  - Only A is considered to be a variable on which we could hypothetically intervene.
- The concept of interaction refers to the joint causal effect of two treatments A and E
  - Interaction involves the counterfactual outcomes $Y^{a,e}$ under a joint intervention
  - Identifying interaction requires exchangeability, positivity, and consistency for both treatments.
  - When treatment E is randomly assigned, then the concepts of interaction and effect modification coincide

WHY DO WE WANT TO ASSESS EFFECT MODIFICATION ?

1. The identification of effect modification may help understand the biological, social, or other mechanisms
2. Understanding Disparities: evaluating the presence of effect modification is helpful to identify the groups of subjects that would benefit the most from an intervention
   - Additive, but not multiplicative, effect modification is the appropriate scale to identify the groups that will benefit the most from intervention (see next slide)
3. If the average causal effect differ between populations with different prevalence of M, it is important for generalizability/transportability:
   - in the presence of an EMM, the average causal effect in this population may not be transportable to other populations with a different distribution of effect modifiers.
   - Not discussed today

METHODS FOR ASSESSING EFFECT MODIFICATION

- Stratified analyses
  - You need to conduct a heterogeneity test
  - Wald test, Cochran Q test…
- Introducing an interaction term in statistical models
  - For additive models (linear models), additive interactions is estimated
  - Not for multiplicative models
- Novel methods for high-dimensional heterogeneous effects

Cochran Q test

$$Cochran's\ Q = \left[\frac{(\beta_1-\beta_P)^2}{VAR(\beta_1)} + \frac{(\beta_2-\beta_P)^2}{VAR(\beta_2)}\right]$$

Where $\beta_1 = \ln(RRstrata_1)$; $\beta_2 = \ln(RRstrata_2)$; VAR is the variance. For the Cochran Q estimation it is necessary to conduct a χ^2 test statistic (with degrees of freedom equal to the number of strata minus 1)

THE IMPORTANCE OF THE SCALE

- Why additive EMM is the most relevant measure in public health ?
- Consider the following example:

table 1

	A=0	A=1
M=0	0.02	0.05
M=1	0.04	0.10

table 2

	A=0	A=1
M=0	0.02	0.05
M=1	0.07	0.10

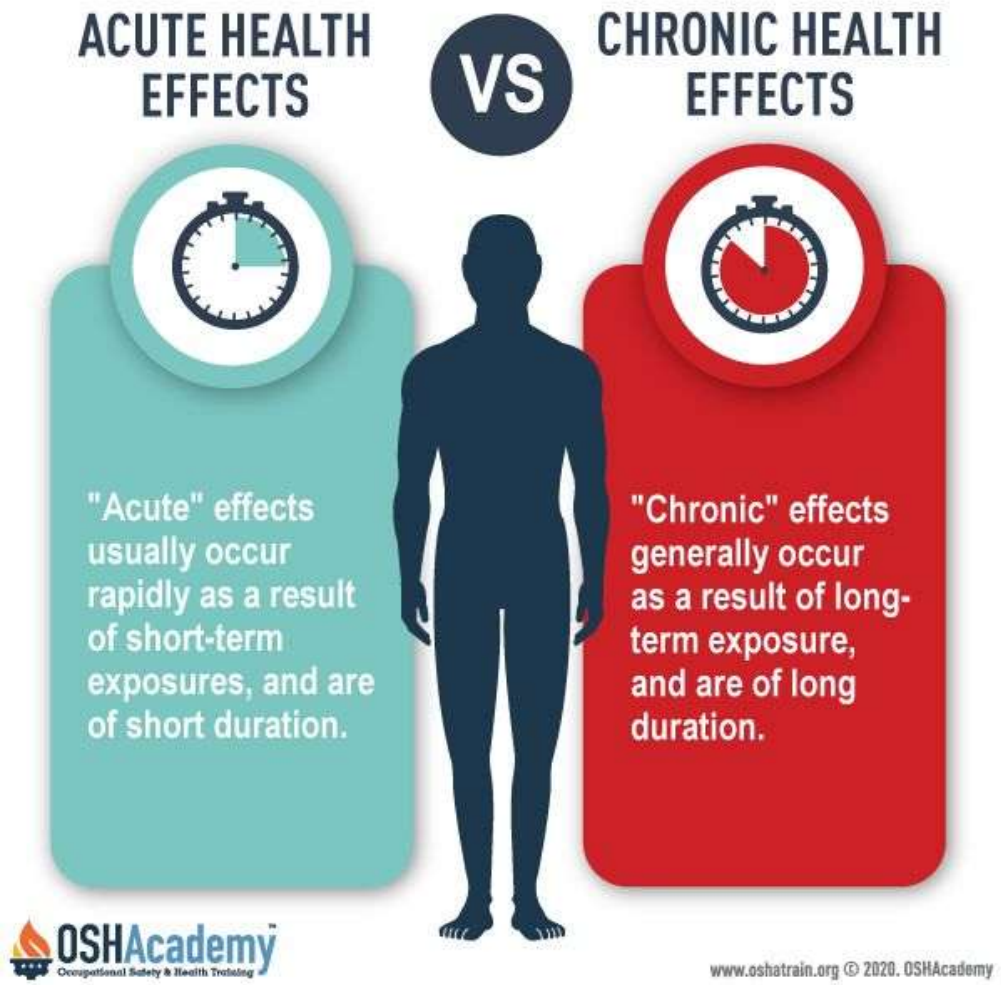
- In table 1:
  - The risk difference for A when M=0 is 0.05 – 0.02 = 0.03
  - The risk difference for A when M=1 is 0.1 – 0.04 = 0.06
- We only have 100 doses of treatment A. What should we do?
- What the is conclusion on the multiplicative scale ?
  - Same risk ratios for M=0 or M=1
- What would be your conclusion from Table 2?

BREAK

STUDY DESIGNS FOR ACUTE HEALTH IMPACTS OF ENVIRONMENTAL EXPOSURES

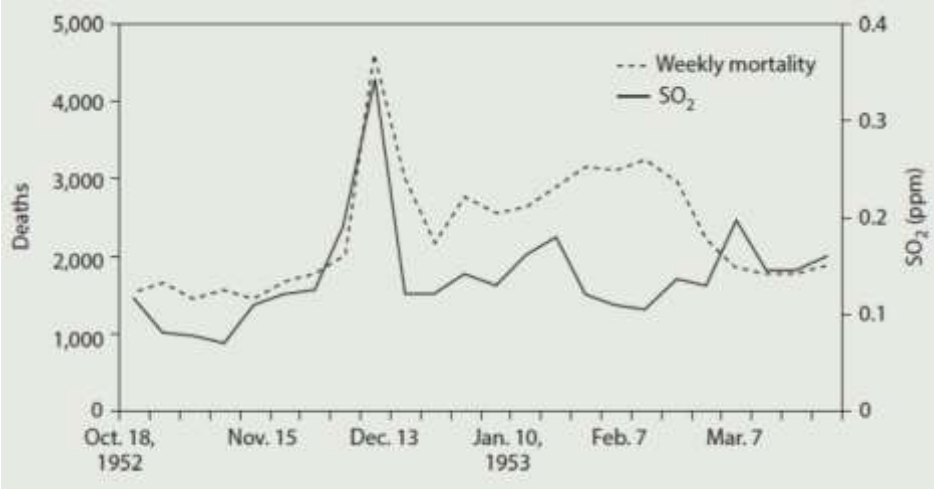
ACUTE EXPOSURE-WHAT DO WE MEAN

- **Acute effects**: Health effects that usually occur rapidly, as a result of short-term exposure
- **Short-term exposure:** in environmental epidemiology, this is usually on time scale of minutes, hours, days, or weeks
  - ❖ Can be a “trigger” or the last step in leading from pathophysiology to disease, or the final component cause leading a susceptible person to experience a specific outcome

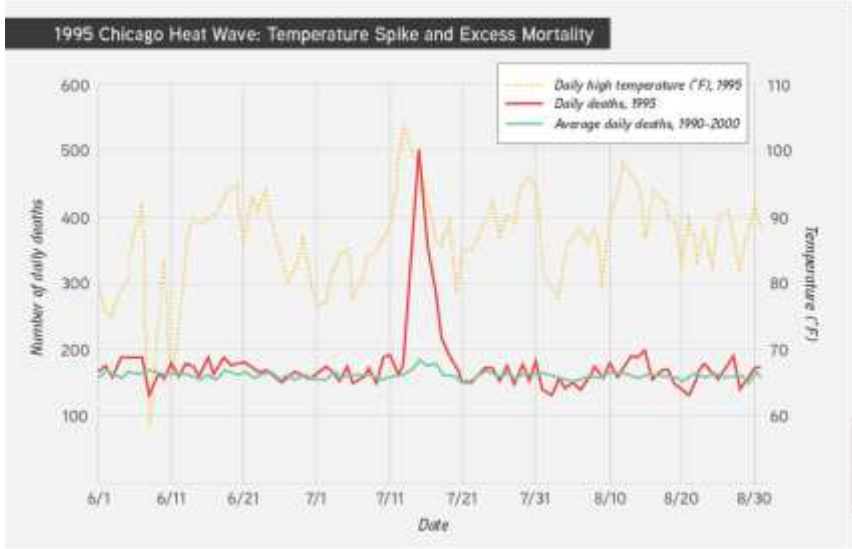


WHY ARE ACUTE ENVIRONMENTAL STRESSORS IMPORTANT TO STUDY

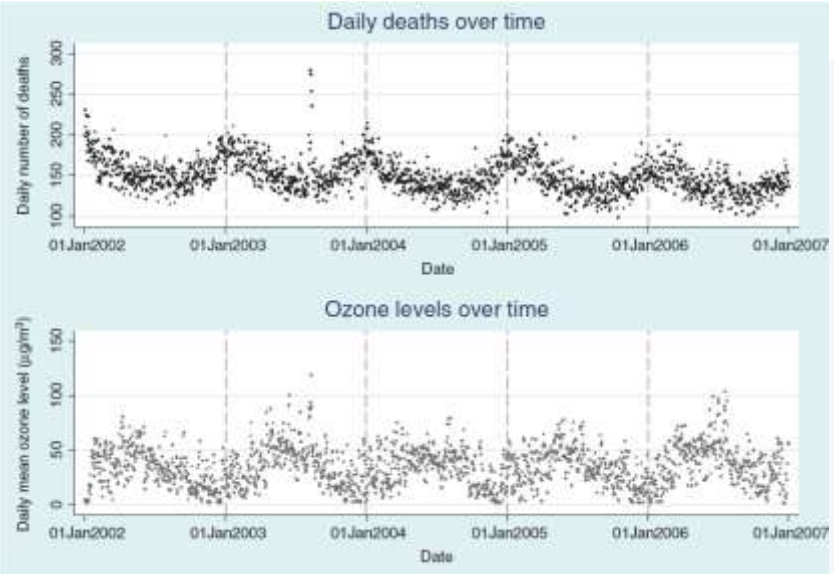
- Acute environmental events which can be studied include (but not limited to):
  - ❖ High air pollution concentrations
  - ❖ Extreme heat or high temperature
  - ❖ Natural disasters such as tornadoes, floods, hurricanes
- Evidence of the effect of acute environmental exposures can be used to inform measures to protect populations during high exposure days
  - ❖ Early warning systems coupled with community-based actions (e.g. Heat action Plans)



(Mortality and SO_2 levels during the 1952 London Smog, adapted from Bell et al., 2001)



HOW DO WE STUDY ACUTE EFFECTS? TIME SERIES DATA!



Example of the outcome and exposure time-series in a time-series study (Figure 1 from Bhaskaran et al. 2013)

Date	Exposure (ex: PM2.5)	Outcome (ex: hosp)	Covariates (ex: temp)
1/1/2016	36.4	116	25.4
1/2/2016	42.3	125	26.9
1/3/2016	40.5	123	20.3
1/4/2016	38.7	119	18.5
…	…	…	…

- Continuous sequence of observations of a population, taken repeatedly over time (normally equal intervals)
- Data for time-series include repeated measures for a given area
  - ❖ Number of health events (ex: daily mortality count)
  - ❖ Exposure levels (ex: concentrations of PM)
  - ❖ Covariate information such as weather variables
- Capitalize on day-to-day (or other time scale) variations in risk factor and in mortality or morbidity counts.

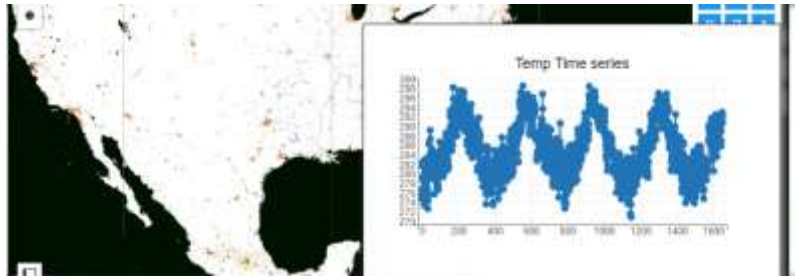
USING TIME SERIES DATA

Data Access

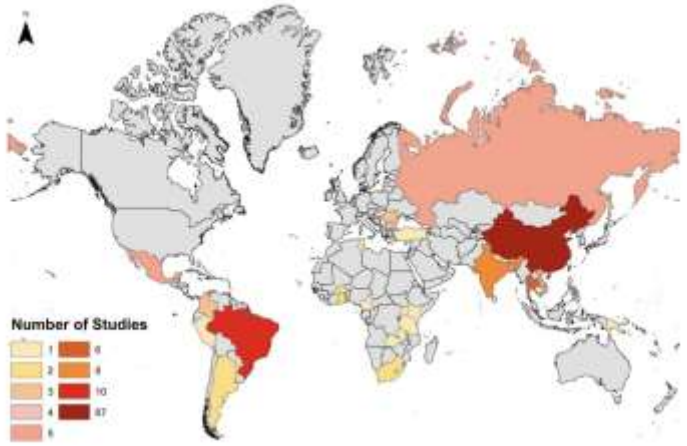
- No need for individual level information
- Confounding by population characteristics is negligible as they are quite stable day to day
- Requires data at a fine temporal scale (ex: daily) for both exposure and outcome

Inference

- Ecological level (usually)
- Large scale studies
- **What about generalizability?**



Satellite products make climate data much more accessible on global scale

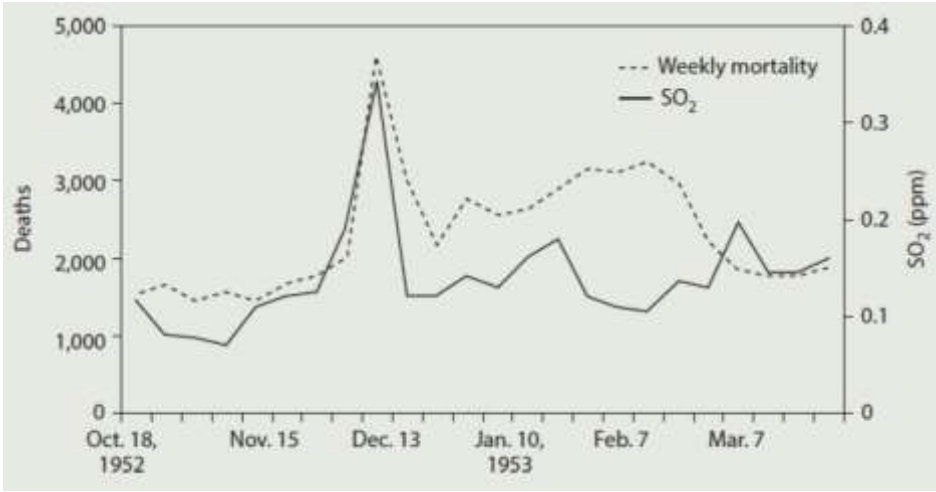


Heat-related in studies in low and middle income countries. (Figure 3 from Green et al., 2019)

TIME SERIES ANALYSIS

Population-based ecological study design to estimate **short-term** health effects of environmental stressors

- Directly models the association between day-to-day variations of exposure and outcome
- Population-based ecological study: we treat the entire study population as exposed and modeled the association between aggregated health outcomes and aggregated exposure.
- Short-term health effects: acute health impacts of environmental stressors, normally within days or weeks.



(Mortality and SO_2 levels during the 1952 London Smog, adapted from Bell et al., 2001)

TIME SERIES ANALYSIS

Use **Poisson regression with over-dispersion** to model the daily count of health outcome:

- **Poisson regression** models count outcome and assumes independence of events after adjusting for confounders
- **Over-dispersion** allows the variance to deviate from mean in the outcome data (quasi-Poisson regression and negative binomial regression)
- Simplified model:

$$\ln\left(E\left(Y_t^c\right)\right) = \text{offset}\left(\ln\left(Pop_{c,t}\right)\right) + \beta^c X_t^c + confounders$$

  - Y_t^c: number of events on day *t* in city *c*
  - $\text{offset}\left(\ln\left(Pop_{c,t}\right)\right)$: population at risk on day *t* in city *c*
  - X_t^c: level of environmental stressor (or heatwave status) on day *t* in city *c*
  - *Confounders:* examples include long-term and seasonal trends in time
  - β^c: city-specific log incidence rate ratio associated with one unit increase in *X* (or comparing heatwave to non-heatwave days)

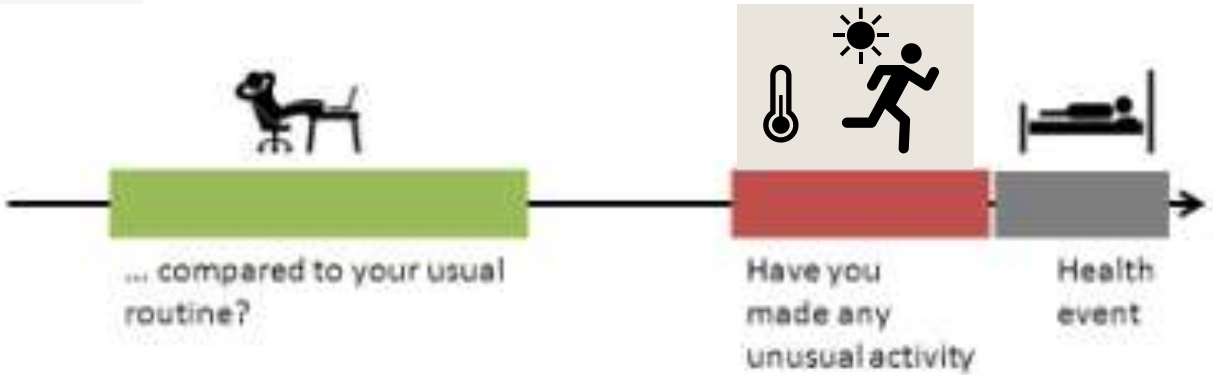
CASE CROSSOVER DESIGN

A combination of crossover trial design and matched case-control design

- Crossover trial design: each patient receives different treatments during different time periods (i.e., patients cross over from one treatment to another during the trial)
- Matched case-control design: select matching controls to cases so that they are similar in confounders to increase efficiency and a matched analysis is generally needed

First proposed by Maclure (1991) to study transient effects on acute health events

- Uses each event or 'case' as its own control
- Identifies 'control' periods in which the event did not occur
- Control for any time-fixed information for by design when using matching method in analysis



CASE CROSSOVER ANALYSIS

Simplified model for conditional logistic regression:

$$logit(P_{i,k}) = \alpha_i + \beta \times X_{i,k} + confounders$$

- $P_{i,k}$ represents the probability of having the event for period *k* in stratum *i*
- α_i represents intercept for stratum *i*
- $X_{i,k}$ represents the value of environmental stressor *X* (or heatwave status) for period *k* within stratum *i*
- β is the log odds ratio of having the event per unit increase in *X* (or comparing heatwave to non-heatwave days)

Extension: treating each period instead of each event as a case

- Conditional logistic regression with weights equal to number of events for each period
- Conditional Poisson regression

OVERVIEW OF CONSIDERATIONS FOR BOTH STUDY DESIGNS

- Estimation of exposure time-series data
  - Time-series analysis: finer or coarser
- Confounding in time-series data
  - Case-crossover: controlled by design + controlled by modeling
  - Time-series analysis: controlled by modeling
- Lag structure
- Pooling of results across geographical units
  - Case-crossover: could skip this step
  - Time-series analysis: multilevel modeling

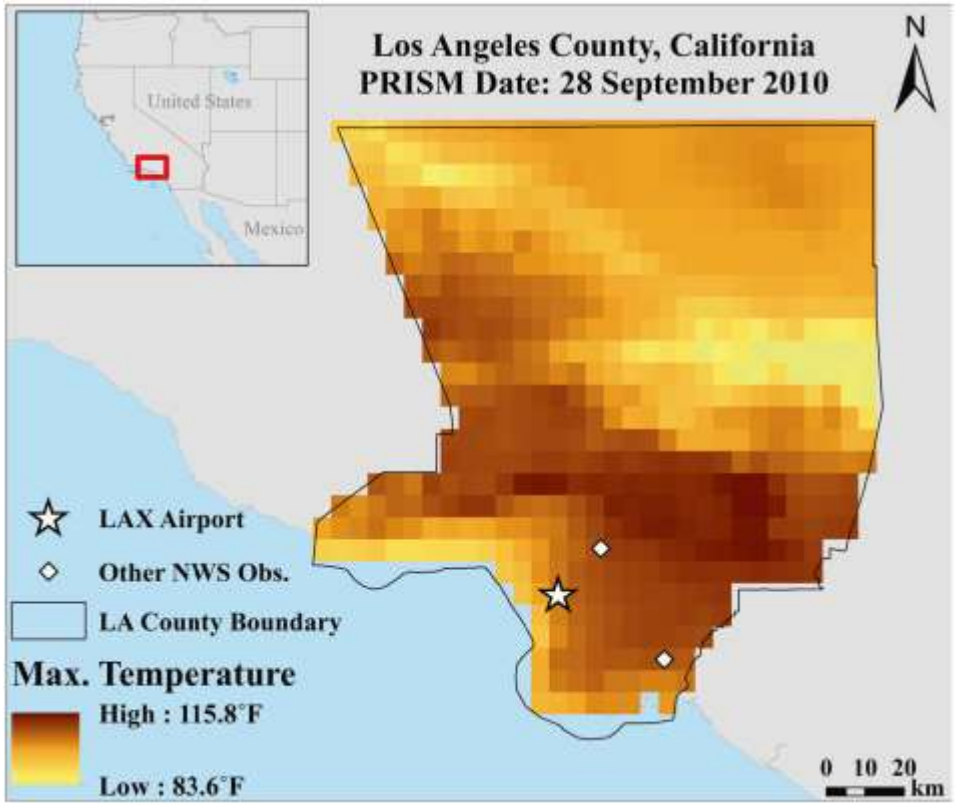
ESTIMATION OF EXPOSURE TIME-SERIES DATA

Exposure data sources:

- Direct measurements from monitoring sites (e.g., weather station)
- Use ensemble model to incorporate indirect info (e.g., satellite and land use data)

In time-series analyses, things to consider in choosing the size of the geographical unit for exposure data aggregation/estimation:

- Maintaining spatial homogeneity of exposure within the unit to approximate population average exposure
  - Population mobility across units
- Having large enough population to satisfy the independence assumption for outcome



Spatial heterogeneity of ambient temperature across Los Angeles county after incorporating indirect info. (Figure 1 from Spangler et al. 2018)

CONFOUNDING IN TIME-SERIES DATA

Temporally stable confounders

- Population composition (e.g., age structure, smoking population)

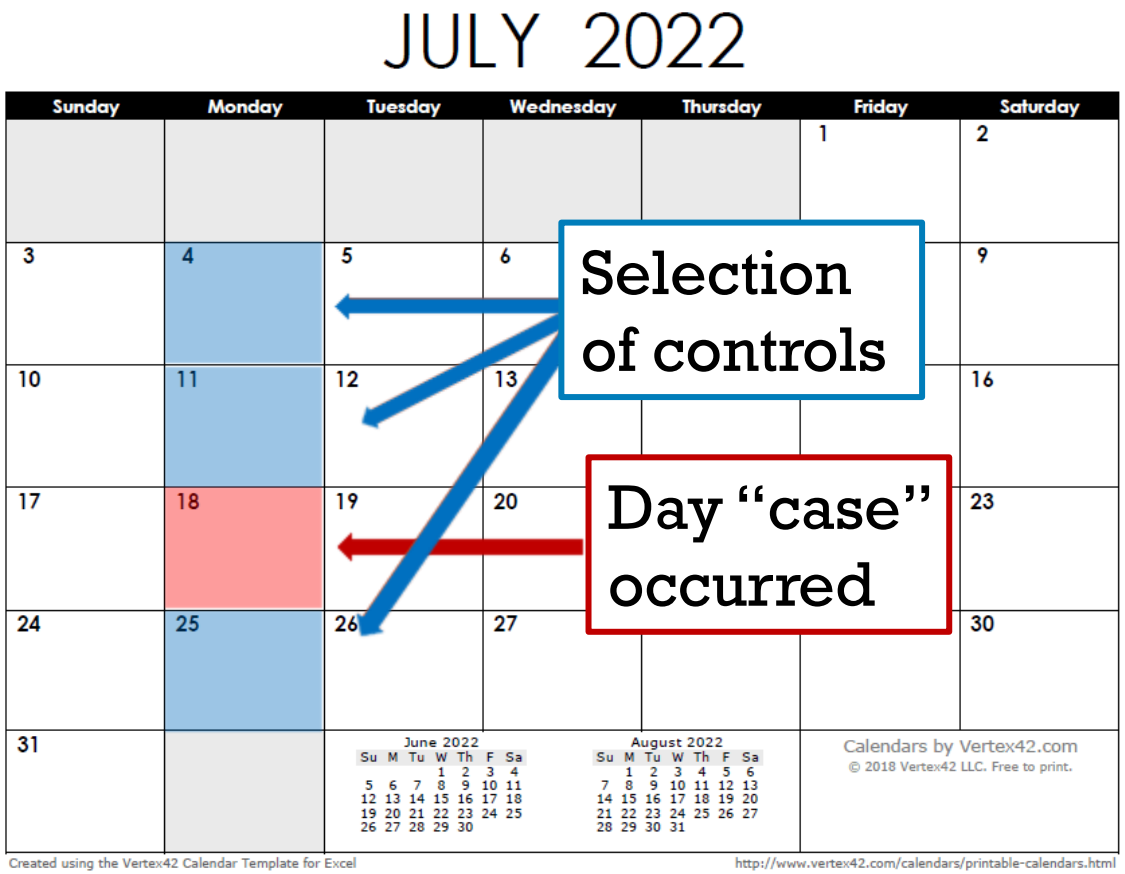
Time-varying confounders

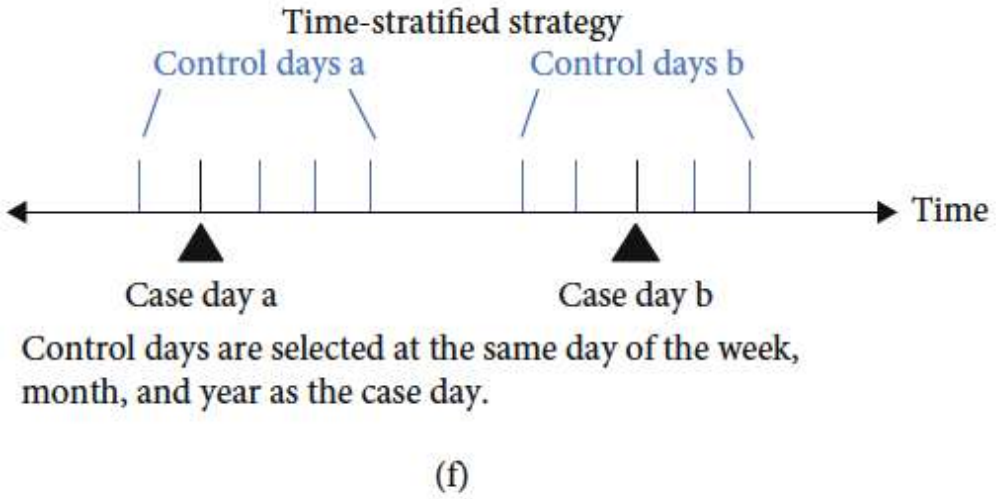
- Day of the week
- Relevant environmental factors (e.g., temperature in studies of air pollution)
- Seasonal trend or long-term pattern (e.g., higher influenza rates and lower temperature in winter)

Types of confounding	Time-series analysis	Case-crossover
Temporally stable confounders	Controlled by design	Controlled by design
Time-varying confounders	Adjusted for in the model	Controlled by design + adjusted for in the model

CASE CROSSOVER DESIGN- SELECTION OF CONTROLS

Time-stratified case crossover design selection of controls





(Figure 1 Wu et al., 2021)

CONTROLLING FOR TIME TRENDS BY DESIGN

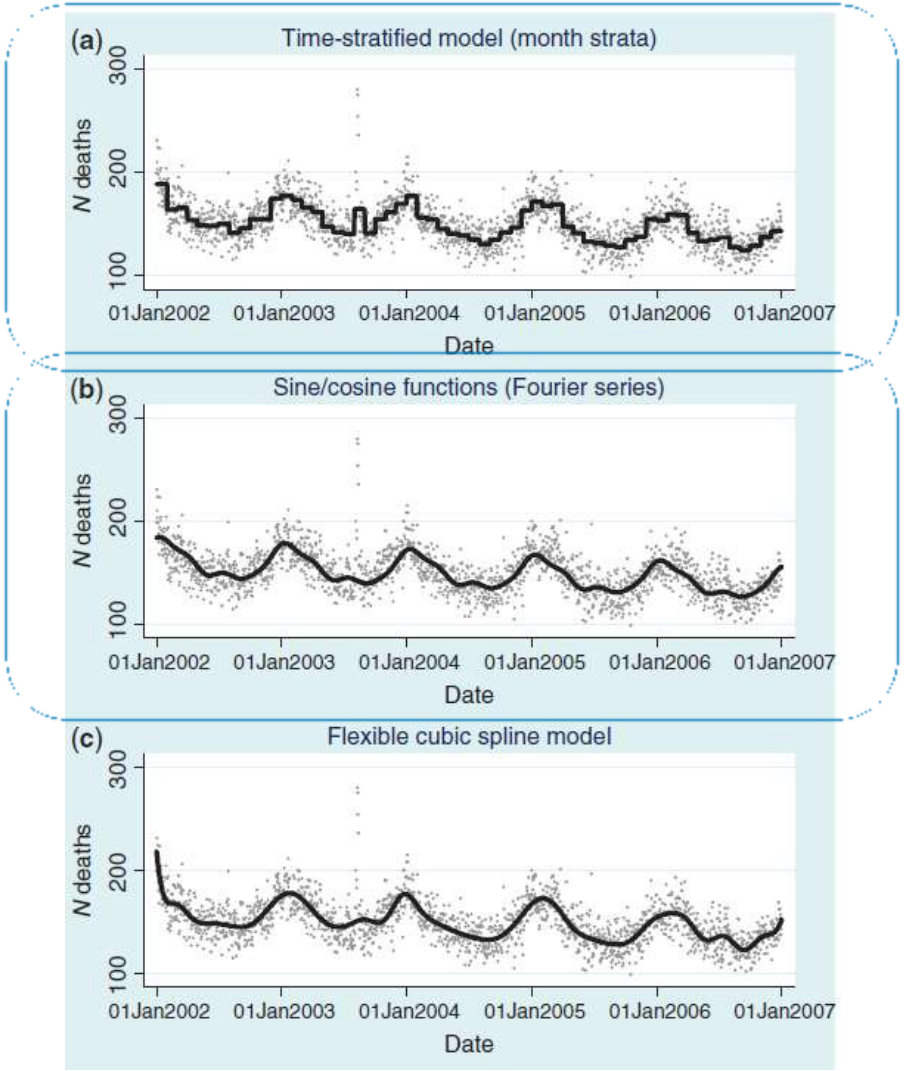
With time-stratified approach, bi-directional selection of controls adjusts for seasonality and time trends by design

- Using the same day of the week to account for behavioral differences between weekdays assuming that activities will be similar on same weekday in a month
- Any time-varying confounder that will differ throughout the month beyond seasonal time trend will still have to be adjusted for in model
  - Ex: temperature in air pollution studies

CONTROL FOR TIME TRENDS IN MODEL

Adjustment for time-varying confounders like seasonal trend, long-term pattern and relevant environmental factors

- Fixed-effect model
  - Split the study period into intervals and estimate each interval separately
  - Cons: abrupt change with large number of parameters
- Periodic function of time
  - Use Fourier terms to summarize the seasonal trend
  - Cons: forced the same trend every year

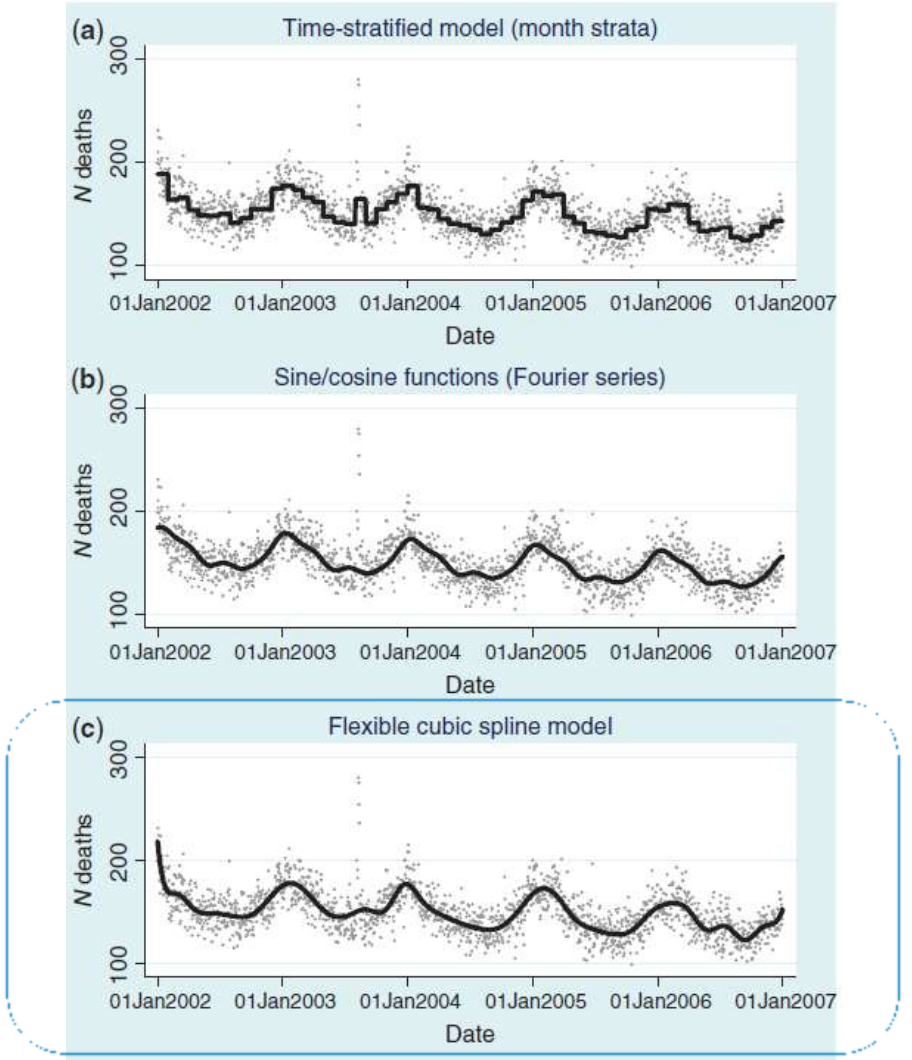


Three ways of modelling long-term pattern (Figure 2 from Bhaskaran et al. 2013)

CONTROL FOR TIME TRENDS IN MODEL

Adjustment for time-varying confounders like seasonal trend, long-term pattern and relevant environmental factors

- Spline function of time: smoothed polynomial curves
  - Degree of freedom: chosen *a prior* based on previous knowledge or chosen based on data



Three ways of modelling long-term pattern (Figure 2 from Bhaskaran et al. 2013)

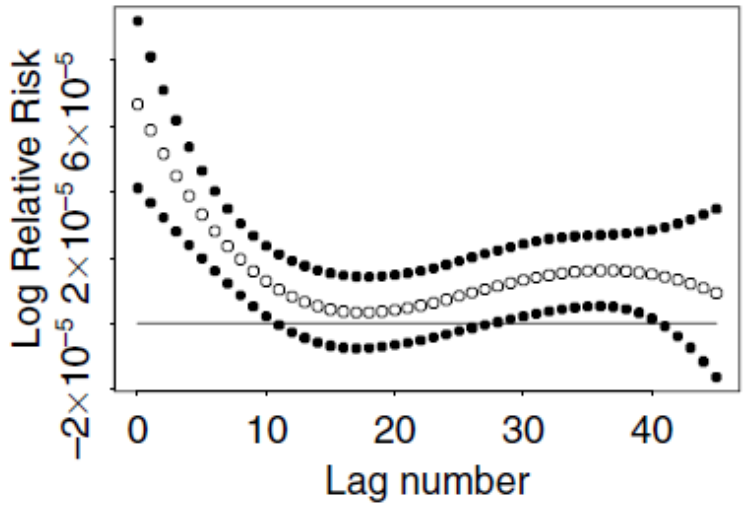
HOW TO CONSIDER LAG STRUCTURES

Existing methods to account for lag structure (applicable to both designs)

- Cumulative effect
  - Moving average of exposure from a few days
- Lagged effect
  - Multiple single-day model with different lag
  - Distributed lag model

Mostly current studies evaluated lag structure up to a week

- Lag structure longer than a week are hard to handle in time-stratified case-crossover design



Association between TSP and mortality over lag structure with penalized spline function (Figure 3 from Zanobetti et al. 2000)

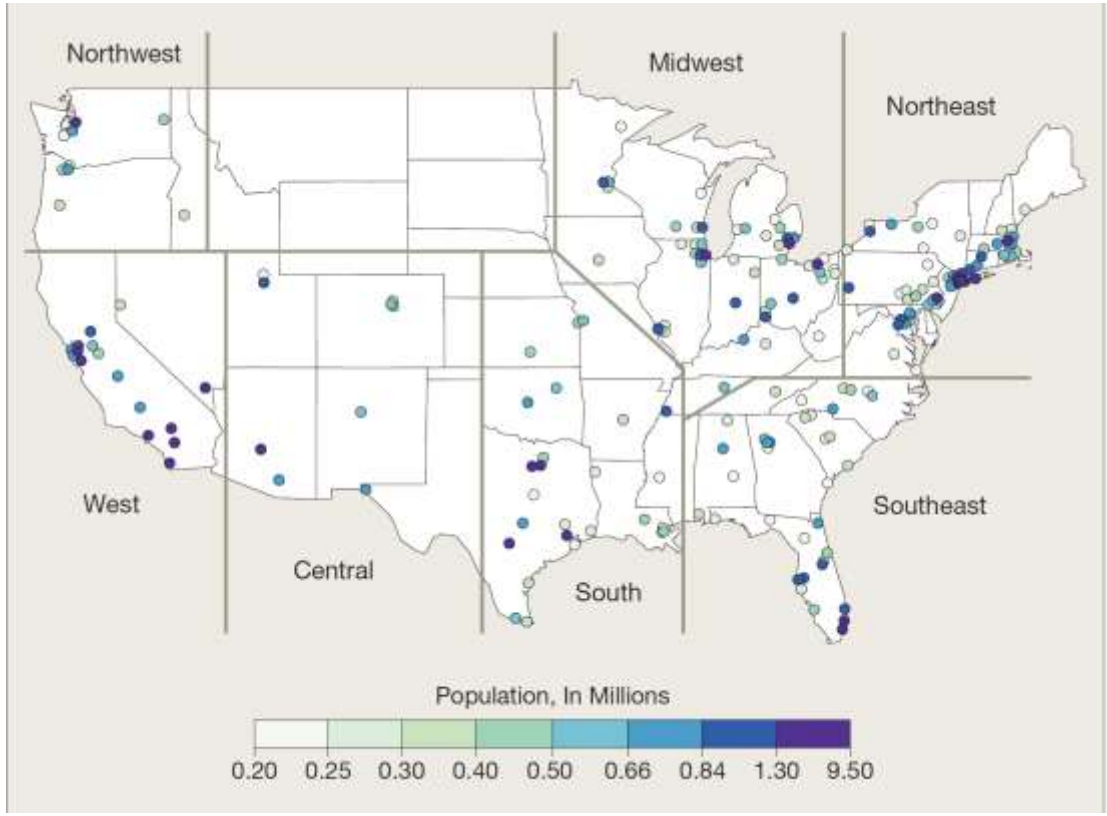
POOLING RESULTS ACROSS GEOGRAPHICAL UNITS

Assuming **an homogeneous concentration-response function (i.e. association/effect) in all geographical units**, we could pool results across them to increase precision

- Time-series analysis: multilevel modeling
  - Bayesian hierarchical modeling
  - Meta-analysis with random intercept for each geographical unit
- Case-crossover design could skip this step
  - Including all risk sets in one model

What if heterogeneity exists?

- Effect modification: more in 2 weeks



Counties included in one study of air pollution (Figure 1 from Dominici et al. 2006)

COMPARISON OF TWO DESIGNS

	Time-series analysis	Case-crossover
Assumptions	Events are independent across days	
	Equal mean and variance in outcome	Same association within each stratum
Analytical method	Poisson model with overdispersion* to estimate rate ratio	Conditional logistic model to estimate odds ratios
Time-varying confounding control	By adjustment in model**	By study design and adjustment in model
Lag structure	Flexible with minimal restriction	Flexible but ambiguity with lagged exposures after 7 days
Result pooling	Each geographical unit needs to be analyzed separately and pooled later	All units could be analyzed together

TWO-WAY FIXED EFFECTS MODELS IN AIR POLLUTION EPIDEMIOLOGY:
A SYSTEMATIC ANALYSIS OF MODEL SENSITIVITY AND RECOMMENDATIONS FOR FUTURE STUDIES

YIQUN MA ET AL.

Two-way fixed effects (TWFE) models

- Common econometric approach used to analyze observational time-series cross-sectional data (aka panel data analysis)
- **What is a fixed effect?**
- Compare each unit to itself over time by including unit fixed effects (also referred to as spatial fixed effects) and adjust for secular trends common across units by including time fixed effects

$$Y_{i,t} = \beta X_{i,t} + \alpha_i + \lambda_t + \varepsilon_{i,t}.$$

unit / spatial fixed effects
(e.g., states or counties)

time fixed effects
(e.g., calendar months or years)

TWFE models have been increasingly applied in air pollution epidemiology

Table 1. Previous epidemiological studies on the health effects of air pollution using two-way fixed effects models

Author (year)	Study period	Study region	Exposure data	Outcome	Fixed effects
Wang et al. (2016)[1]	2004-2009	New Jersey, U.S.	Area-weighted annual mean $PM_{2.5}$	All-cause mortality	census tract + year
Renzi et al. (2019)[2]	2006-2012	Latium, Italy	Area-weighted annual mean PM_{10}	All-cause mortality	municipality + year
Leogrande et al. (2019)[39]	2008-2014	Taranto, Italy	Population-weighted annual mean industrial PM_{10}	Cause-specific mortality	district + year + age class + district^year + age class^year
Yu et al. (2020)[40]	1990-2013	Queensland, Australia	Population-weighted annual mean $PM_{2.5}$	All-cause and cause-specific mortality	postcode area + year
Han et al. (2021)[41]	2000, 2010	China	Population-weighted annual mean $PM_{2.5}$	All-cause mortality	county + year
Yu et al. (2022)[3]	2010-2016	Brazil	Area-weighted wildfire-related annual mean $PM_{2.5}$	Cancer mortality	municipality + year
Yu et al. (2022)[42]	2010-2018	Brazil	Population-weighted annual mean $PM_{2.5}$	Loss of life expectancy	municipality + year
Nyadanu et al. (2022)[43]	2012-2019	Ghana	Area-weighted annual mean $PM_{2.5}$	Stillbirths	district + year
Fan et al. (2023)[44]	2015-2020	Jiangsu, China	Population-weighted annual mean $PM_{2.5}$	Cancer mortality	spatial unit + year
Guo et al. (2023)[10]	2000-2015	Brazil	Daily mean $PM_{2.5}$ value extracted by municipality centers	Hospitalizations	panel + day
Wang et al. (2023)[45]	2002-2016	U.S.	Area-weighted annual mean $PM_{2.5}$, NO_2, and O_3	Myocardial infarction hospitalizations	zip code + year + age group
Heft-Neal et al. (2023)[4]	2006-2017	California, U.S.	Population-weighted daily mean $PM_{2.5}$	Cause-specific emergency department visits	zip code + year^season + county^month + day-of-week
Li et al. (2023)[46]	2016-2020	Jiangsu, China	Population-weighted annual mean $PM_{2.5}$ and $PM_{2.5}$-bound polyaromatic hydrocarbon	Cancer mortality	spatial unit + year
Yu et al. (2024)[47]	1987-2000	U.S.	Station-based annual mean $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO	All-cause and cause-specific mortality	spatial unit + year
Li et al (2024)[48]	2015-2020	Jiangsu, China	10-year population-weighted $PM_{2.5}$ constituents	Cancer mortality	spatial unit + year

Note: We used the symbol "^" to denote the interaction between fixed effects. This table is not a systematic review and may not provide a comprehensive summary of the literature.

Advantages of TWFE models

- flexibly applied to study the health impacts of either short- or long-term exposure to air pollution
- offers a feasible alternative when some other study designs are impractical or impossible to implement
- inherently account for unmeasured confounders that are fixed over time or space when appropriate fixed effects are included

Developing a TWFE model involves various choices in model specifications

- **Aggregation of environmental data**
  - E.g., area-weighted or population-weighted
- **Covariates**
  - Temperature, precipitation, …
  - Linear or non-linear terms
- **Combination of fixed effects**
  - County + year, county + month, …

- All these decisions can be reasonable, but they involve different assumptions and may lead to nontrivial differences in research findings.
- It remains unclear how sensitive TWFE models and related inferences are to different modeling choices and how to transparently select a final model from a wide range of options.

Study aims

Image created by ChatGPT 4

As an illustrative case study, we investigated associations between monthly wildfire smoke $PM_{2.5}$ concentrations and all-cause mortality in 3,084 counties in the contiguous U.S., 2006–2019

This study aims to

a) systematically assess the sensitivity of TWFE models to specification choices
b) propose a comprehensive decision-making framework

Study design overview

identify sensitive aspects of TWFE models

1. Define the reference model

- a TWFE model with a simple setting to serve as the reference model (Model 0)

2. Run models with alternative specifications

- compare the reference model to a series of modified models (Models 1-1 to 4-48)

demonstrate a final model selection process

3. Propose and test candidate models

- a set of candidate models based on a priori knowledge about the research question

4. Select the final model

- select the final model based on permutation tests and a principle of parsimony

The reference model

We applied a TWFE model with a quasi-Poisson regression to estimate the association between monthly wildfire smoke $PM_{2.5}$ exposure and mortality.

$$\ln\left[E\left(Mortality_{c,m,y}\right)\right] = \beta_1 SmokePM_{c,m,y} + \beta_2 Temperature_{c,m,y} + \beta_3 Precipitation_{c,m,y} + offset\left[\ln\left(Population_{c,y}\right)\right] + \alpha_c + \gamma_y + \varepsilon_{c,m,y}.$$

- $Mortality_{c,m,y}$: *all-cause deaths in county c, month m, year y*
- $SmokePM_{c,m,y}$: *population-weighted mean smoke PM2.5 concentration in county c, month m, year y*
- $Temperature_{c,m,y}$: *population-weighted mean air temperature in county c, month m, year y*
- $Precipitation_{c,m,y}$: *population-weighted total precipitation in county c, month m, year y*
- $\ln\left(Population_{c,y}\right)$: *population offset*
- α_c: county fixed effects, which controls for characteristics that are unique to each county but do not change over time
- γ_y: year fixed effects, which accounts for characteristics that are unique to each year but do not vary across counties
- $\varepsilon_{c,m,y}$: error term

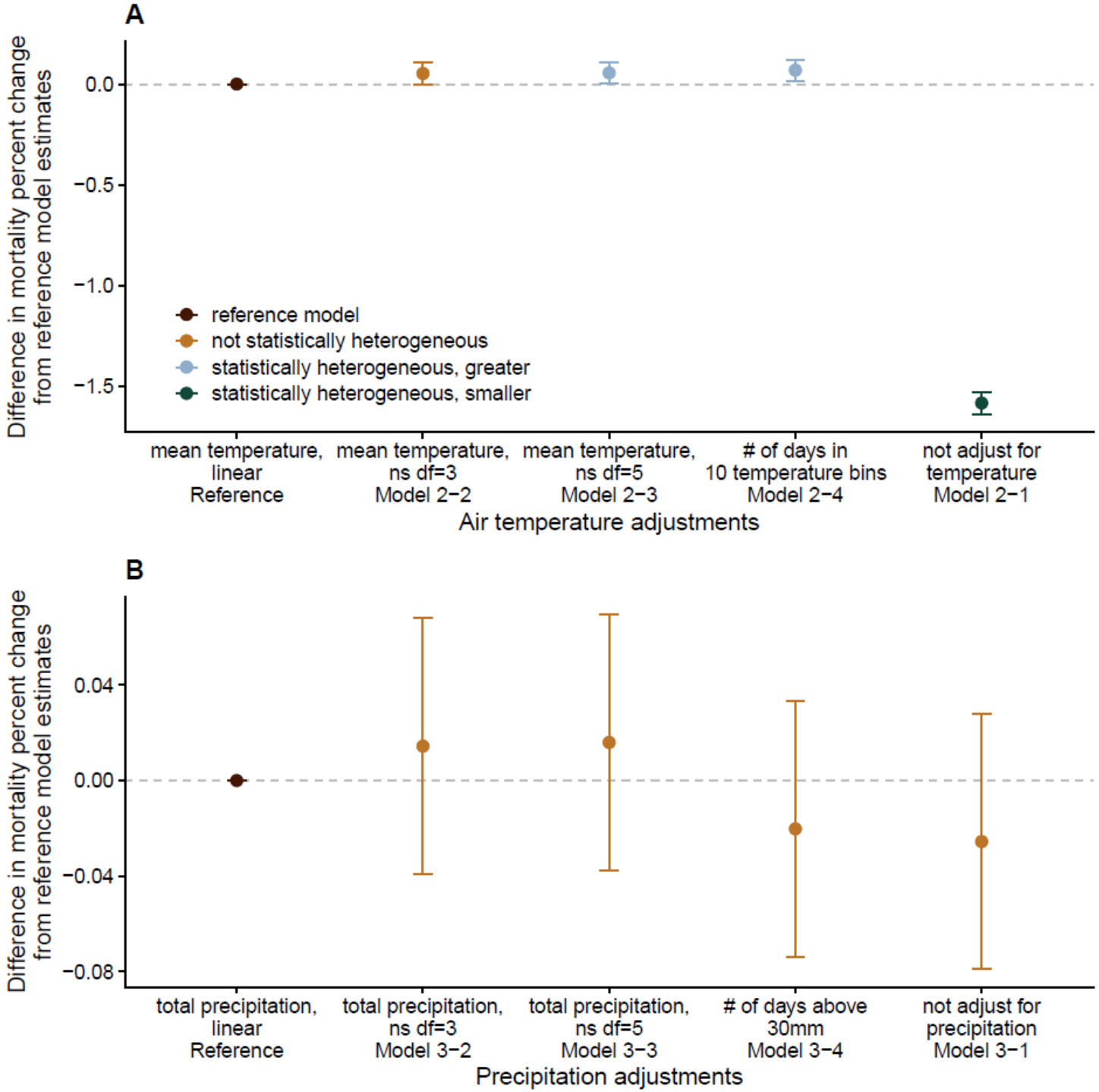
Models with alternative specifications

Each modified model is an alternative specification of the reference model, differing in only one aspect to isolate the effect of that particular change.

Table 2. List of models with different specifications

Category of specifications	Data aggregation	Temperature adjustment	Precipitation adjustment	Fixed effects	Model #
Reference	population-weighted smoke $PM_{2.5}$, temperature, and precipitation	mean temperature, linear	total precipitation, linear	county + year	Reference (Model 0)
Data aggregation	1. population-weighted smoke $PM_{2.5}$ and temperature; area-weighted precipitation 2. population-weighted smoke $PM_{2.5}$; area-weighted temperature and precipitation 3. population-weighted smoke $PM_{2.5}$ and precipitation; area-weighted temperature 4. area-weighted smoke $PM_{2.5}$; population-weighted temperature and precipitation 5. area-weighted smoke $PM_{2.5}$ and precipitation; population-weighted temperature 6. area-weighted smoke $PM_{2.5}$ and temperature; population-weighted precipitation 7. area-weighted smoke $PM_{2.5}$, temperature, and precipitation	mean temperature, linear	total precipitation, linear	county + year	Models 1-1 to 1-7
Temperature adjustment	population-weighted smoke $PM_{2.5}$, temperature, and precipitation	1. mean temperature, ns df=3 2. mean temperature, ns df=5 3. # of days in 10 bins 4. not adjust for temperature	total precipitation, linear	county + year	Models 2-1 to 2-4
Precipitation adjustment	population-weighted smoke $PM_{2.5}$, temperature, and precipitation	mean temperature, linear	1. total precipitation, ns df=3 2. total precipitation, ns df=5 3. # of days above 30 mm 4. not adjust for precipitation	county + year	Models 3-1 to 3-4
Fixed effects combinations	population-weighted smoke $PM_{2.5}$, temperature, and precipitation	mean temperature, linear	total precipitation, linear	See **Methods S1**	Models 4-1 to 4-48

Note: SE, standard error; IID, independent and identically distributed; ns, natural cubic spline; df, degree of freedom.



Temperature adjustment

- Compared to the reference model, modeling air temperature as a non-linear term resulted in higher estimates
- Not adjusting for air temperature in the model led to a substantial smaller estimate.

Precipitation adjustment

- Different methods of adjusting for total precipitation, or not adjusting for precipitation, did not result in statistically heterogeneous estimates.

Figure 2. Differences in estimates between models with alternative air temperature (*A*) and precipitation (*B*) adjustments and the reference

Combination of fixed effects

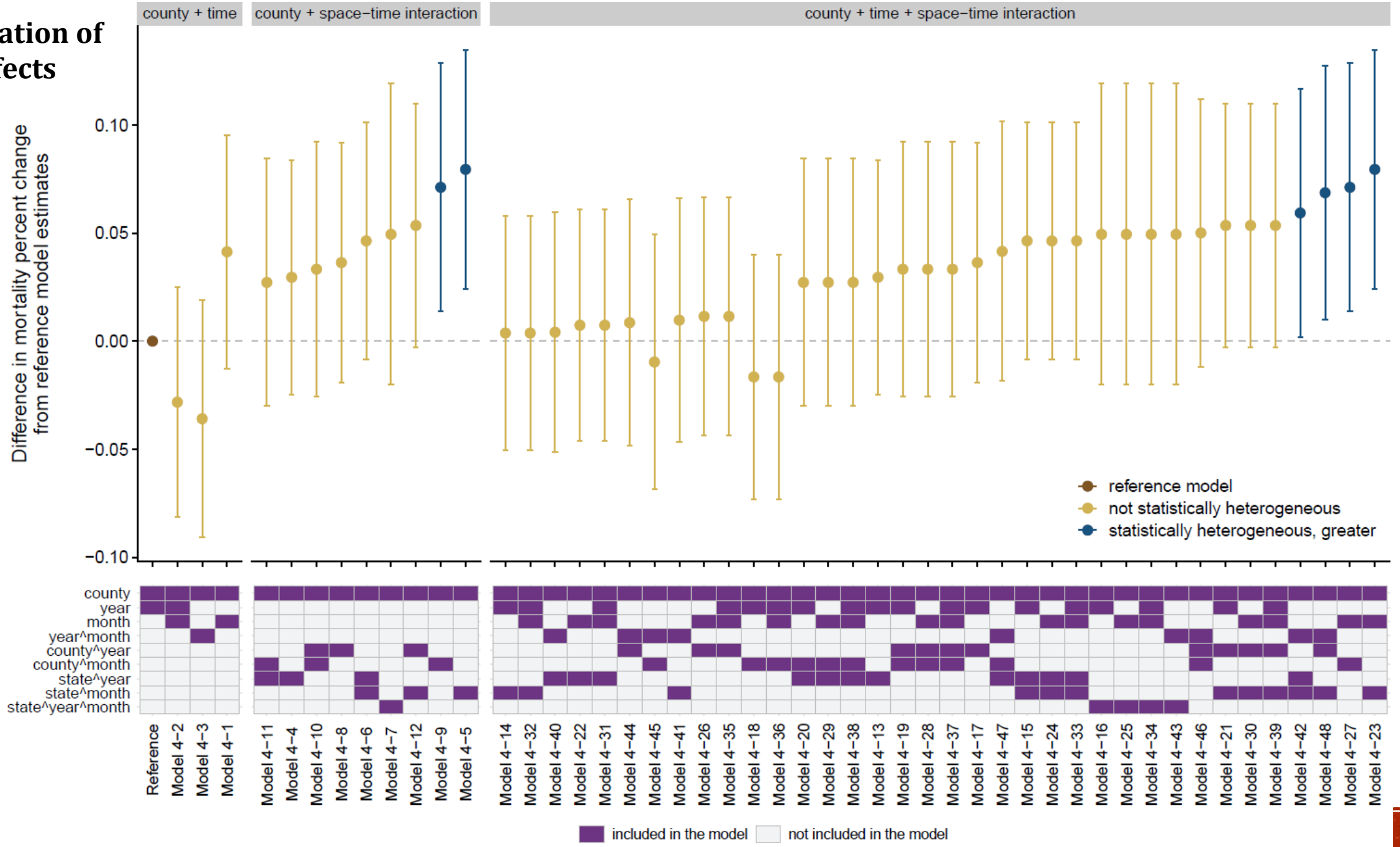


Figure 3. Differences in estimates between models with alternative combinations of fixed effects and the reference model.

List of candidate models

Some decisions can be made by researchers' knowledge:

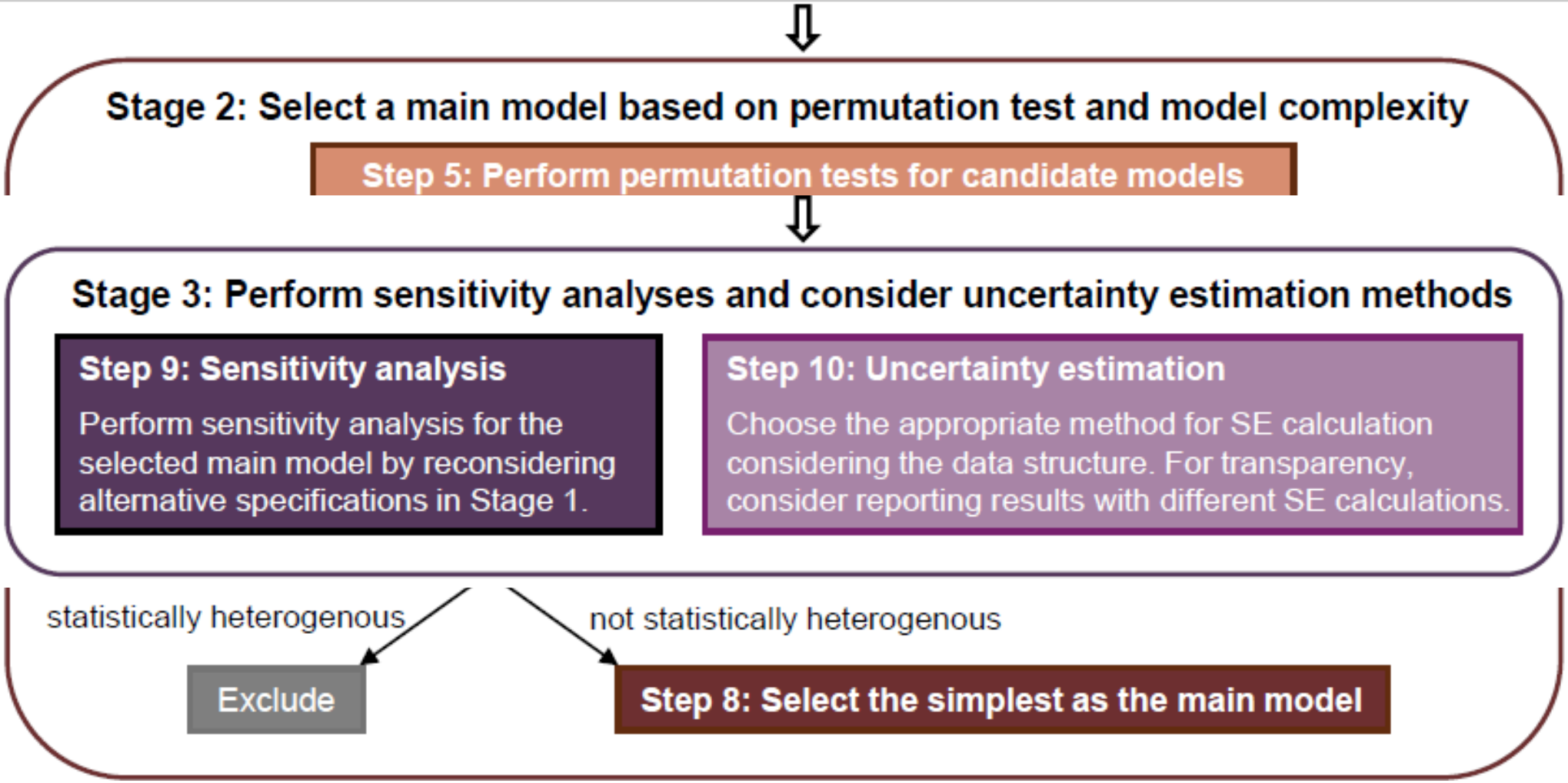
- Population- vs. area-weighted:
  - No meaningful difference in estimates
  - Population-weighted data is preferable
- Covariate adjustments:
  - Failing to adjust for air temperature or neglecting its non-linear nature led to underestimations
  - Air temperature (ns, 5 dfs)
  - Model is not sensitive to adjustment for precipitation
  - Precipitation (linear term)

- Fixed effects:
  - we incorporated various plausible combinations of fixed effects in these candidate models

Table 3. List of candidate models

Candidate model #[1]	Fixed effects
Model 5-1	county + year
Model 5-2	county + year + month
Model 5-3	county + year^month
Model 5-4	county + year + month + state^year + state^month
Model 5-5	county + year^month + state^year + state^month
Model 5-6	county + state^year^month
Model 5-7	county + year + county^month
Model 5-8	county + year^month + state^year + county^month
Model 5-9	county + year^month + county^year + state^month
Model 5-10	county + county^year + county^month
Model 5-11	county + year + month + county^year + county^month
Model 5-12	county + year^month + county^year + county^month

Proposed decision-making framework



Summary

1. TWFE model estimates were particularly sensitive to air temperature adjustments and the combination of fixed effects (and specific interactions terms) employed.
2. We propose a comprehensive framework to guide decision-making in the development of TWFE models, considering the range of modeling options available.
3. This framework has potential applications in future epidemiological studies across diverse settings.

The R code to replicate these analyses is available at the following link:
https://github.com/YiqunnMaa/TWFE_sensitivity

Image created by

Acknowledgements

Yiqun Ma, Chen Chen and Lara Schwarz and the Climate Epidemiology Group

The Workshop Organizing Committee

THANK YOU

YIM022@UCSD.EDU
TBENMARHNIA@UCSD.EDU