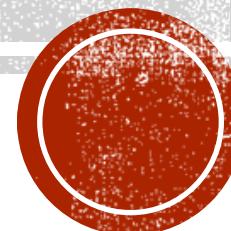


ECO-HEALTH WORKSHOP: INTRO TO CLIMATE EPI METHODS PART I



28 October 2025

Tarik Benmarhnia & Yiqun Ma

University of California, San Diego

UC Center for Climate Health and Equity

UNIVERSITY
OF
CALIFORNIA

Center for Climate,
Health and Equity

OUTLINE

- A framework for climate change epidemiology
- A general review of causal inference
- Study designs for acute health impacts of climate-sensitive exposures
 - Time series modelling and case crossover designs
 - Two-way Fixed Effect (TWFE) Analyses for air pollution epidemiology
 - Natural experiments and interrupted time series analyses

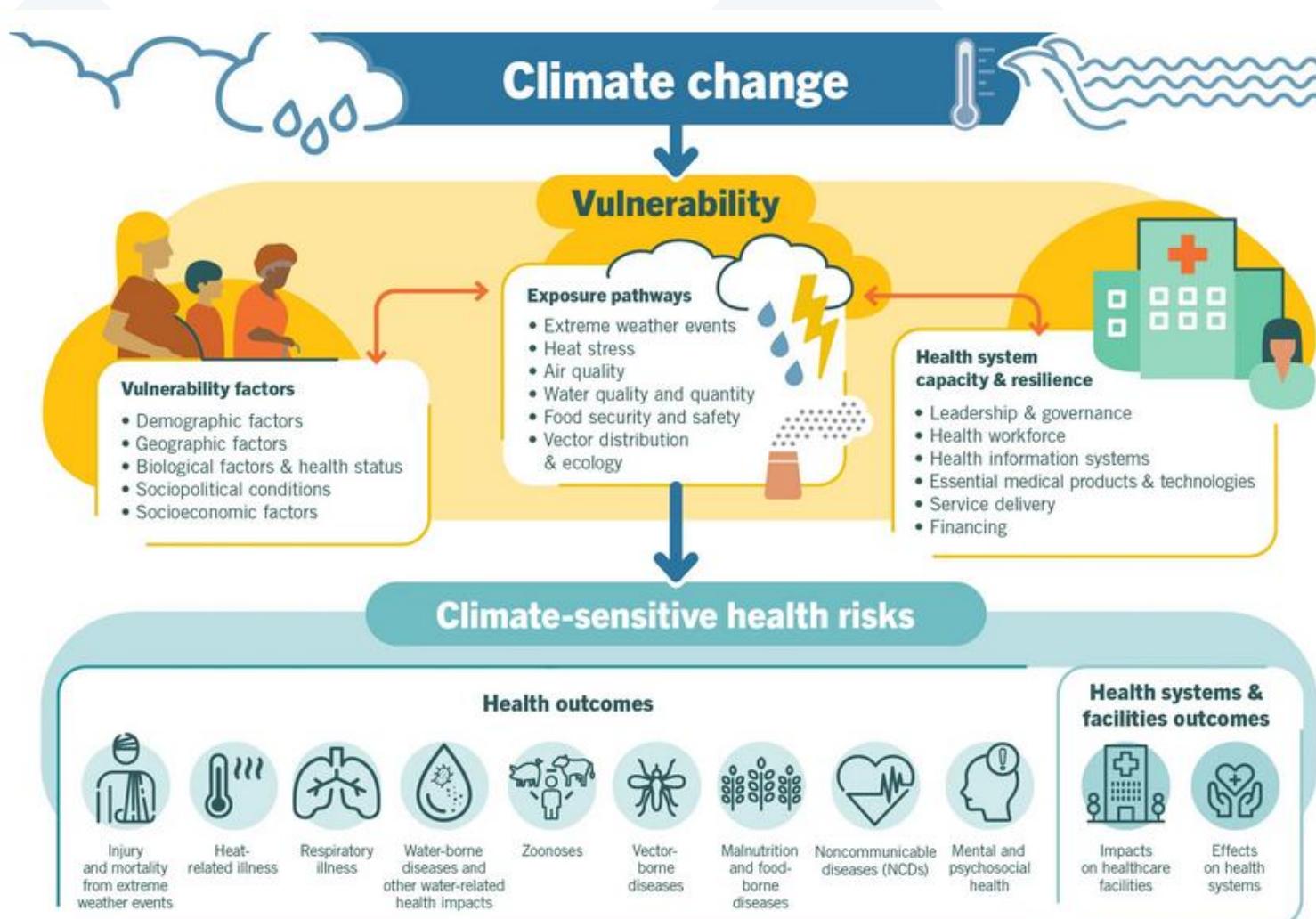


GITHUB LINK

- [https://github.com/benmarhnia-lab/EcoHealth Climate and Health Research Workshop](https://github.com/benmarhnia-lab/EcoHealth_Climate_and_Health_Research_Workshop)



Climate Change and Health: from impacts to adaptation



CLIMATE CHANGE AS A CONTEXT

1. That exacerbates existing weather conditions
2. That creates unprecedented situations
3. An opportunity for mitigation policies with co-benefits on public health

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:

- Preterm Birth and wildfire 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	<ul style="list-style-type: none">• Eligibility criteria• Features (symptoms, clinical parameters ...)	<ul style="list-style-type: none">• Eligibility criteria• Outcome (diagnosis of stroke over the next year)• Inputs (age, blood pressure, history of stroke, diabetes at baseline)	<ul style="list-style-type: none">• 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 ...

ASSUMING AND ASKING GOOD CAUSAL QUESTIONS

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)

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

Miguel A. Hernán, MD, DrPH



A PRAGMATIC APPROACH TO CAUSAL INFERENCE

- We want to infer that X causes Y
 - We can summarize this relationship using a measure of association/correlation/dose-response...
- We cannot demonstrate that X is a cause of Y, instead, we eliminate alternative explanations
- We can for example consider 6 possible explanations for which X is correlated with Y
 - Only one possible explanation is of interest: X causes Y

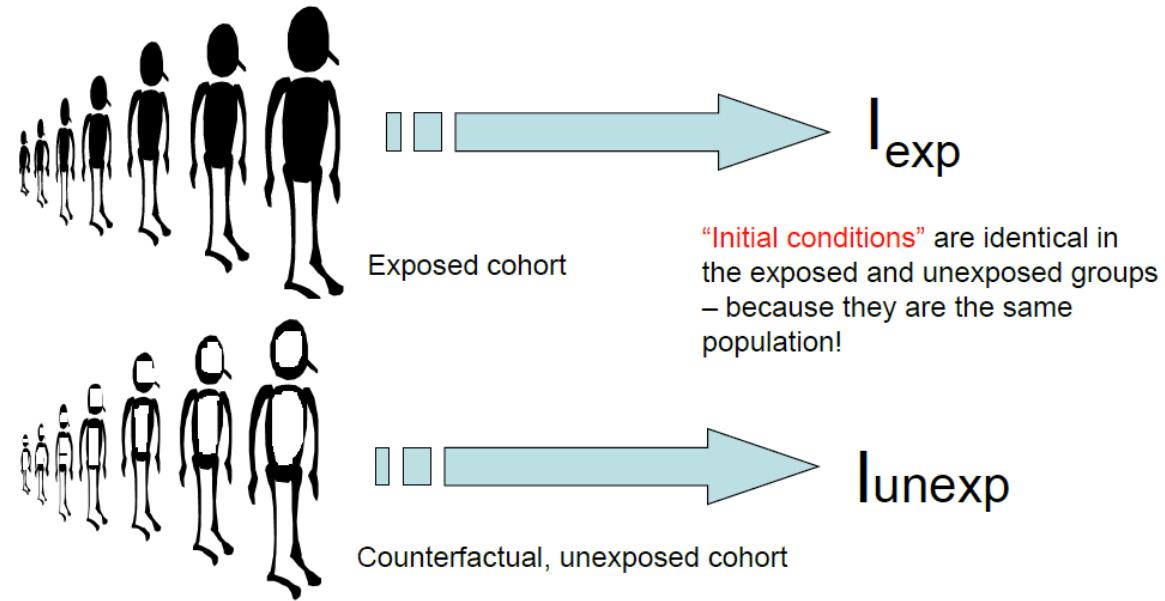


THE COUNTERFACTUAL MODEL

- 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



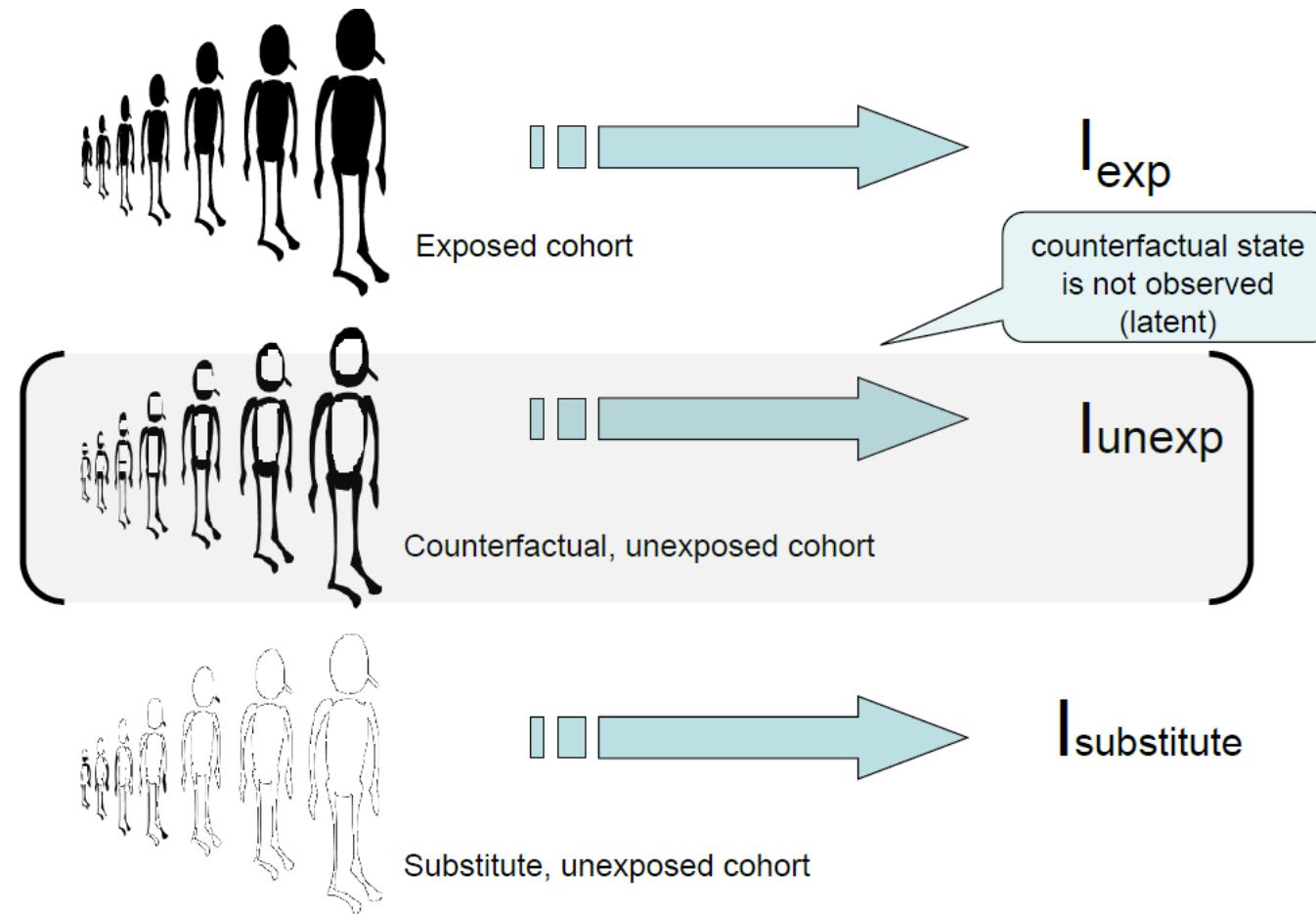
"A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period"

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

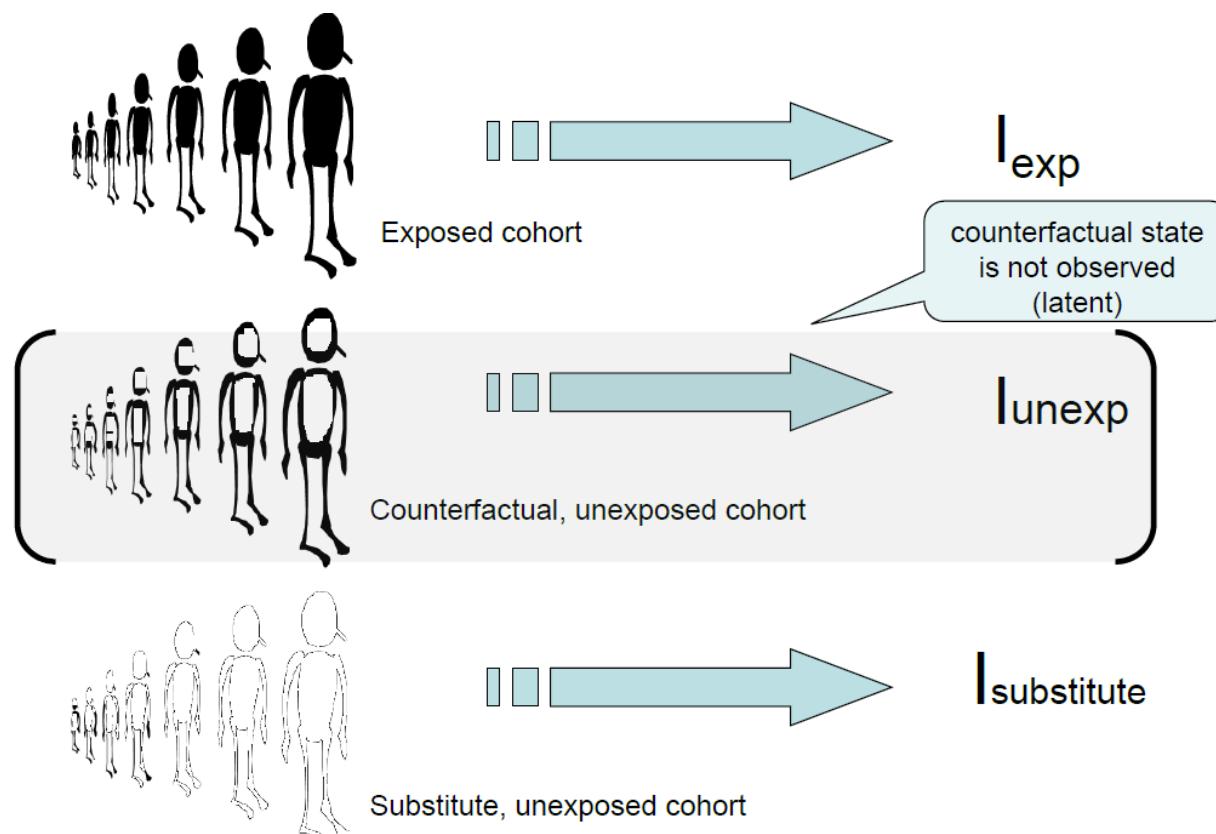
The **fundamental problem** of causal inference is that, by definition, the counterfactual outcome is impossible to observe.



COUNTERFACTUALS IN PRACTICE



FINDING A SUBSTITUTE FOR THE COUNTERFACTUAL GROUP



Some Notations

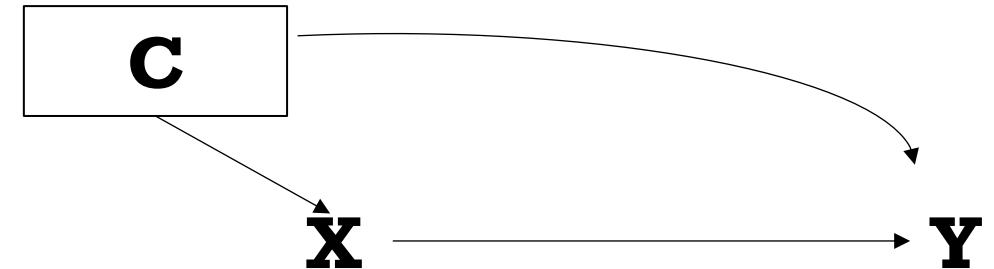
- (i) $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] = 0$
- (ii) $\frac{\Pr[Y^{a=1} = 1]}{\Pr[Y^{a=0} = 1]} = 1$
- (iii) $\frac{\Pr[Y^{a=1} = 1]/\Pr[Y^{a=1} = 0]}{\Pr[Y^{a=0} = 1]/\Pr[Y^{a=0} = 0]} = 1$

MAIN ASSUMPTIONS IN THE COUNTERFACTUAL FRAMEWORK

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



CONFOUNDING BIAS



- Some definitions:

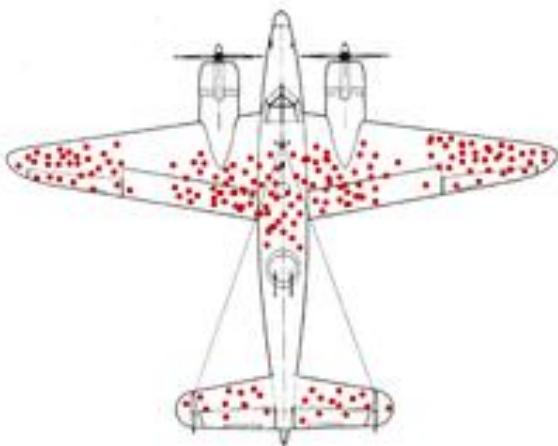
- Confounding is a type of bias that arises from specific causal structures, typically a shared cause of X and Y
- If **conditioning** upon a set of covariates Z will render the association between X and Y unconfounded, then Z is a sufficient set of covariates for estimating the X-Y relationship
- Back-door path to be closed

- **Conditioning** on a covariate:

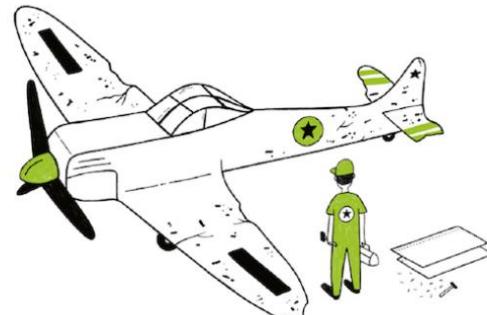
- Restriction
- Sample selection (can be unknown or involuntary)
- Matching
- Stratification
- Regression Adjustment
- Standardization, g-formula
- IP weighting



SELECTION BIAS



SURVIVORSHIP BIAS



The example of Basketball players

X is height

Y is speed

Z is Basketball skills

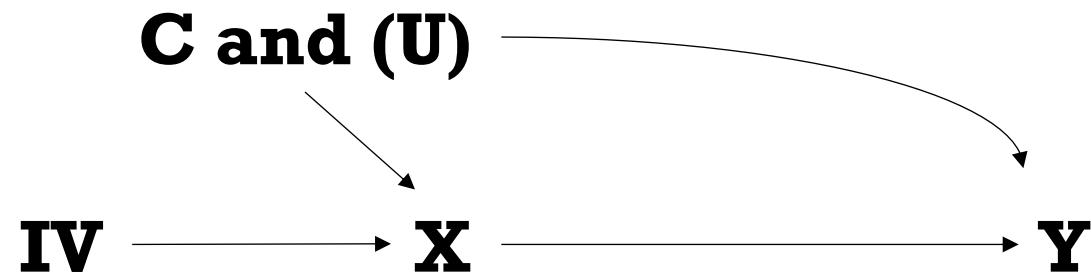
- Short people are not necessarily fast, nor are tall people;
- However, if you look only at professional basketball players, you may find that the short ones are very fast
- By **restricting to pro basketball players**, you have conditioned on a common effect of height and speed, and within this stratum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

The birthweight paradox



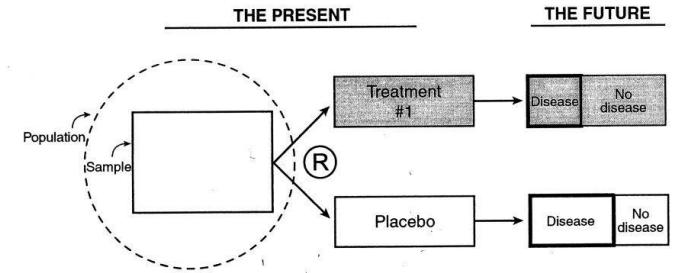
THE IDEA BEHIND RANDOMIZATION

- 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



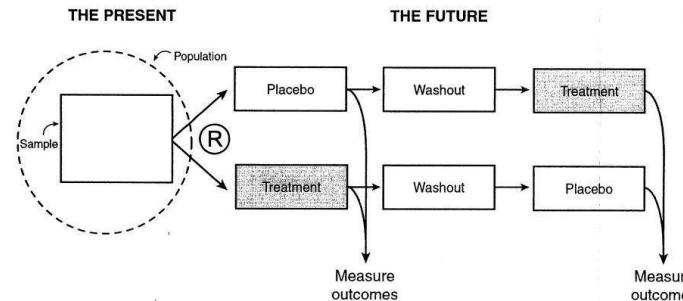
DIFFERENT TYPES OF RCTS

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



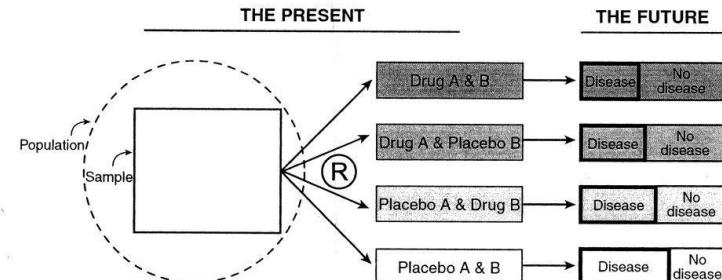
Two-arm, parallel design

■ 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.



Planned cross-over 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.

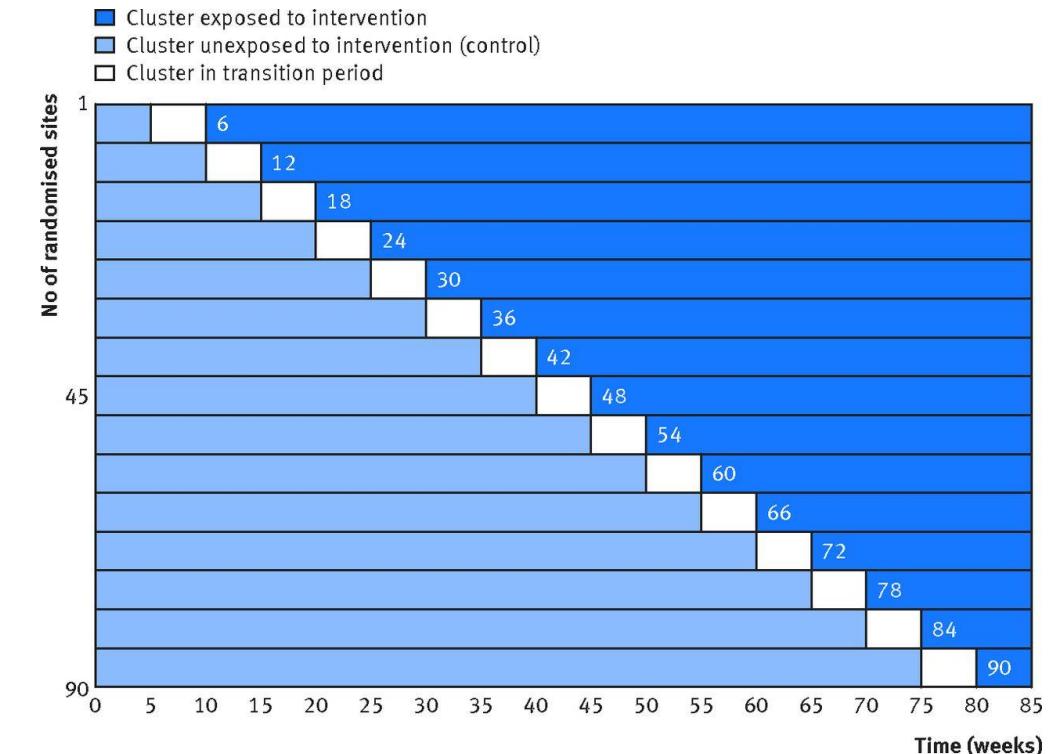


Factorial design

■ 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.

DIFFERENT TYPES OF RCTS

- Two-arm, parallel design
- Planned cross-over design
- Factorial design
- Cluster randomized trials
- **Stepped Wedge Designs**
 - Based on the timing of the intervention



Hemming, Karla, et al. "The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting." *Bmj* 350 (2015): h391.

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}



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

BMJ VOLUME 327 20-27 DECEMBER 2003 bmj.com

Hazardous journeys

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

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. Labrecque¹ · Sonja A. Swanson¹

Practice of Epidemiology

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

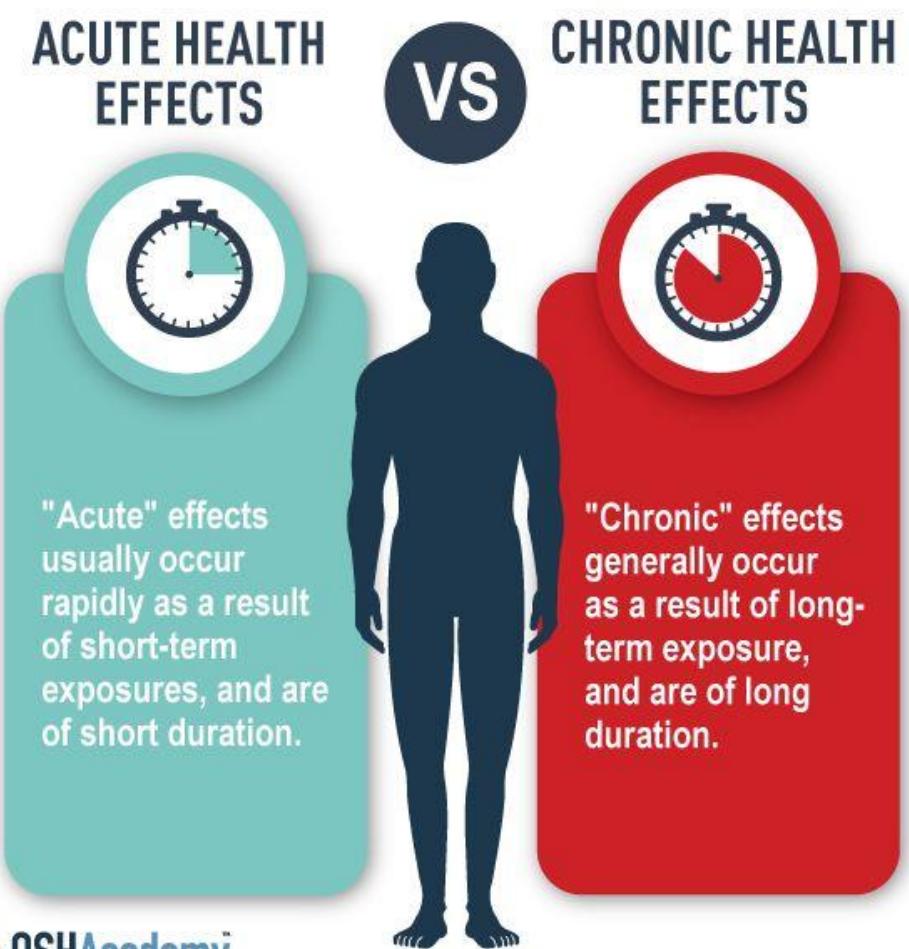
Miguel A. Hernán* and James M. Robins

STUDY DESIGNS FOR ACUTE HEALTH IMPACTS OF ENVIRONMENTAL EXPOSURES



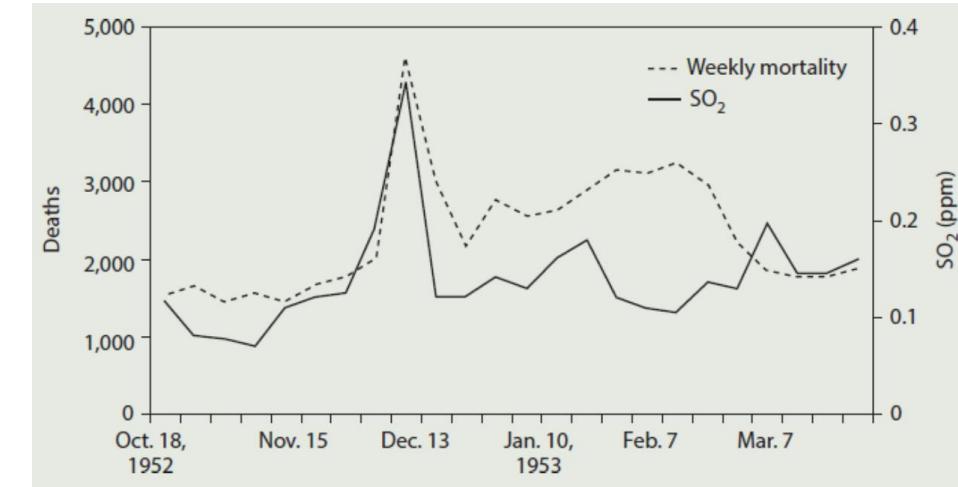
ACUTE EXPOSURE-WHAT DO WE MEAN

- **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

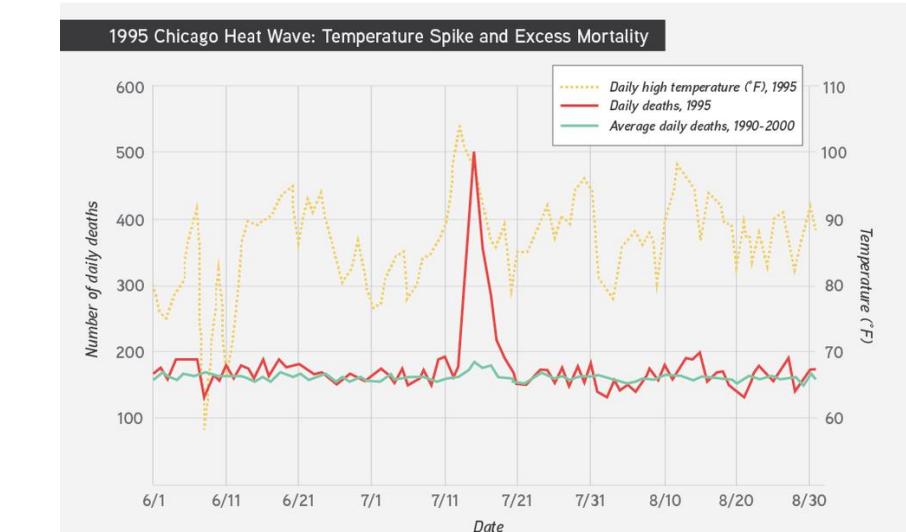


WHY ARE ACUTE ENVIRONMENTAL STRESSORS IMPORTANT TO STUDY

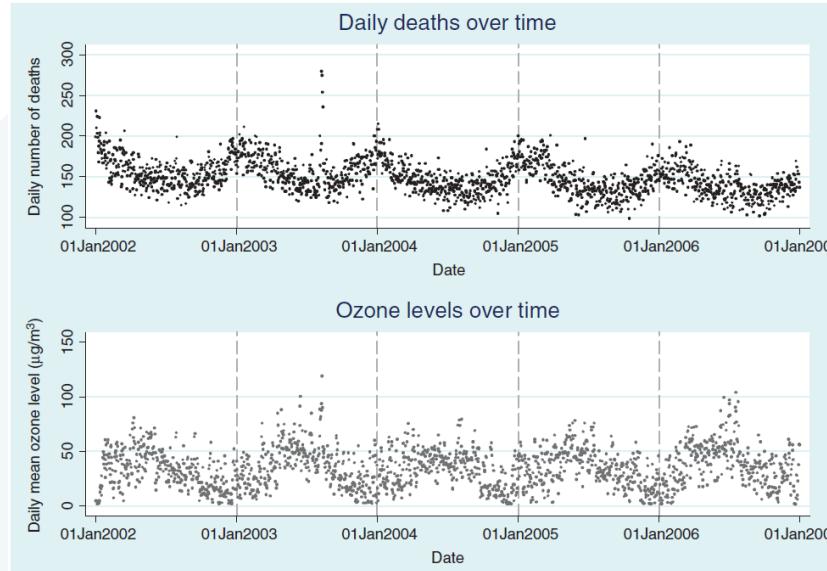
- Acute environmental events which can be studied include (but not limited to):
 - ❖ High air pollution concentrations
 - ❖ Extreme heat or high temperature
 - ❖ Natural disasters such as floods, hurricanes
- Evidence of the effect of acute environmental exposures can be used to inform measures to protect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



(Mortality and SO₂ 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)

- 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.

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
...



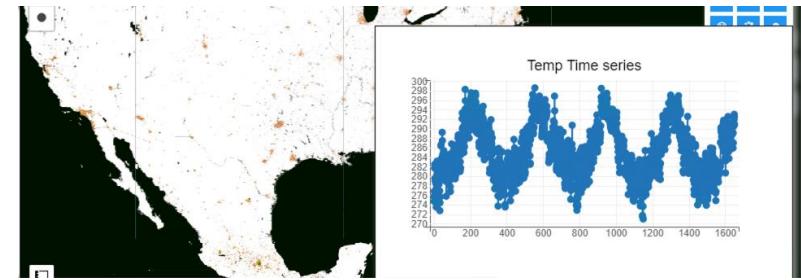
USING TIME SERIES DATA

Data Access

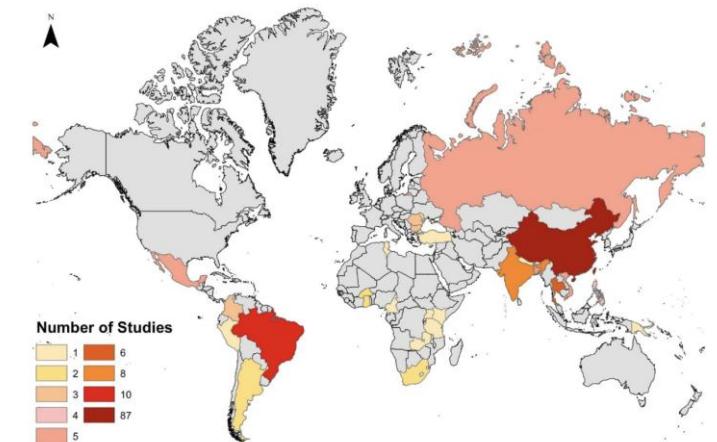
- ❖ No need for individual level information
- ❖ Confounding by population characteristics is negligible as they are quite stable day to day
- ❖ Requires data at a fine temporal scale (ex: daily) for both exposure and outcome

Inference

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



Satellite products make climate data much more accessible on global scale

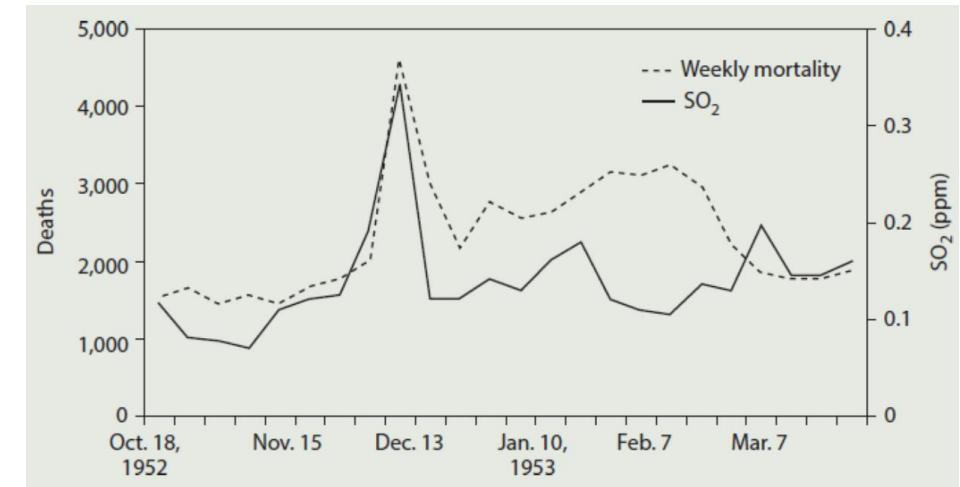


Heat-related in studies in low- and middle-income countries.
(Green et al., 2019)

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, typically within days or weeks.



(Mortality and SO₂ 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:

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

- ❖ Y_t^c : number of events on day t in city c
- ❖ $\text{offset}(\ln(Pop_{c,t}))$: 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:

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

- ❖ $P_{i,k}$ 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)

Alternative model specifications:

- ❖ 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



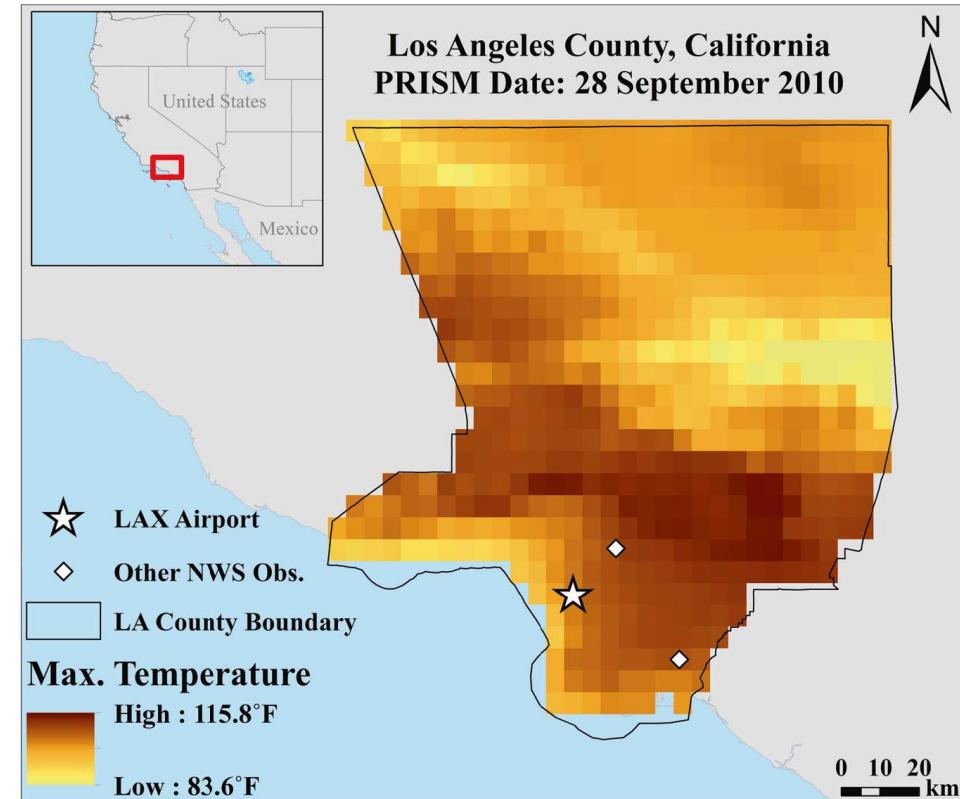
ESTIMATION OF EXPOSURE IN TIME-SERIES DATA

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

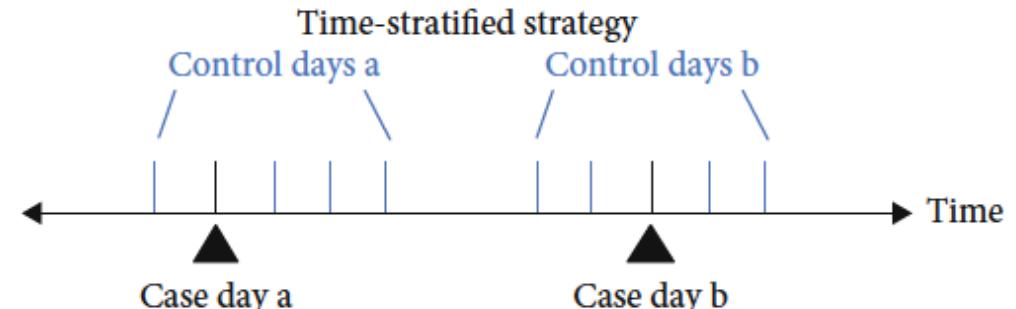
Time-stratified case crossover design selection of controls

JULY 2022

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
					1	2
3	4	5	6			9
10	11	12	13			16
17	18	19	20			23
24	25	26	27			30
31		June 2022	August 2022	Calendars by Vertex42.com © 2018 Vertex42 LLC. Free to print.		
		Su M Tu W Th F Sa	Su M Tu W Th F Sa			
		1 2 3 4	3 4 5 6			
		5 6 7 8 9 10 11	7 8 9 10 11 12 13			
		12 13 14 15 16 17 18	14 15 16 17 18 19 20			
		19 20 21 22 23 24 25	21 22 23 24 25 26 27			
		26 27 28 29 30	28 29 30 31			

Selection
of controls

Day “case”
occurred



Control days are selected at the same day of the week, month, and year as the case day.

(f)

(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

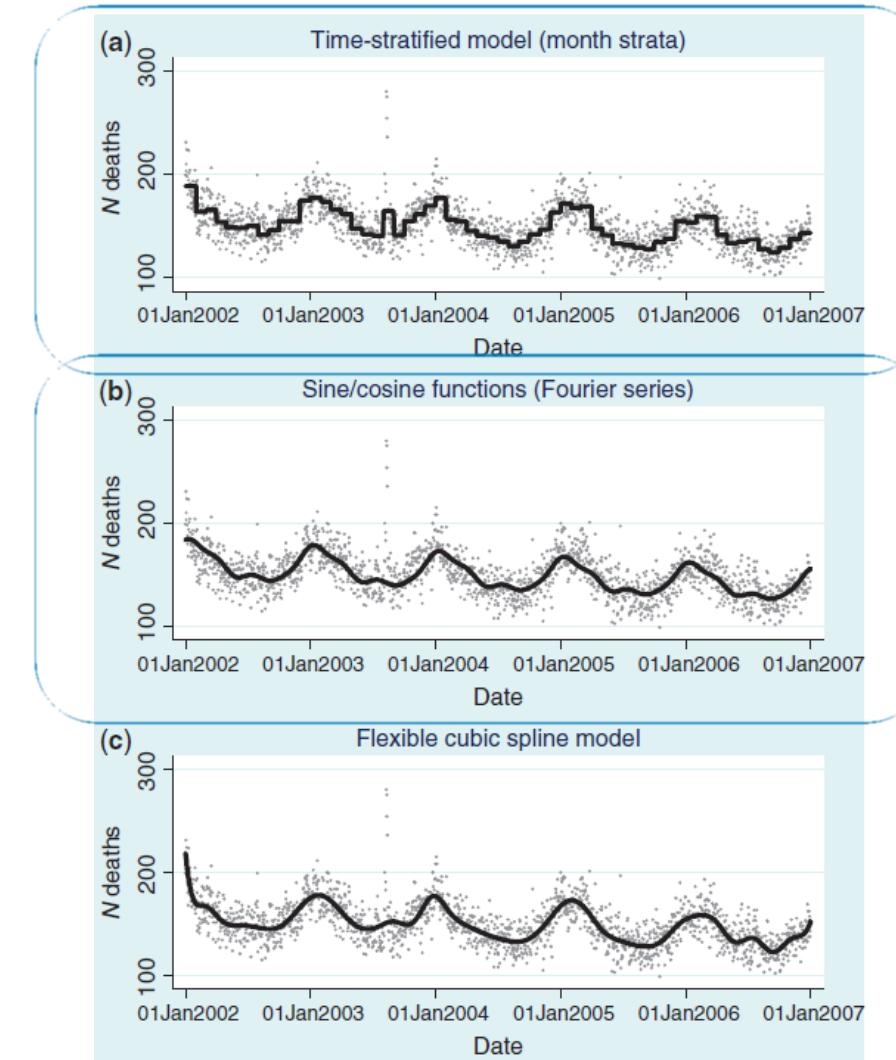
- ❖ Using the same day of the week to account for behavioral differences between weekdays assuming that activities will be similar on same weekday in a month
- ❖ Any time-varying confounder that will differ throughout the month beyond seasonal time trend will still have to be adjusted for in model
 -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

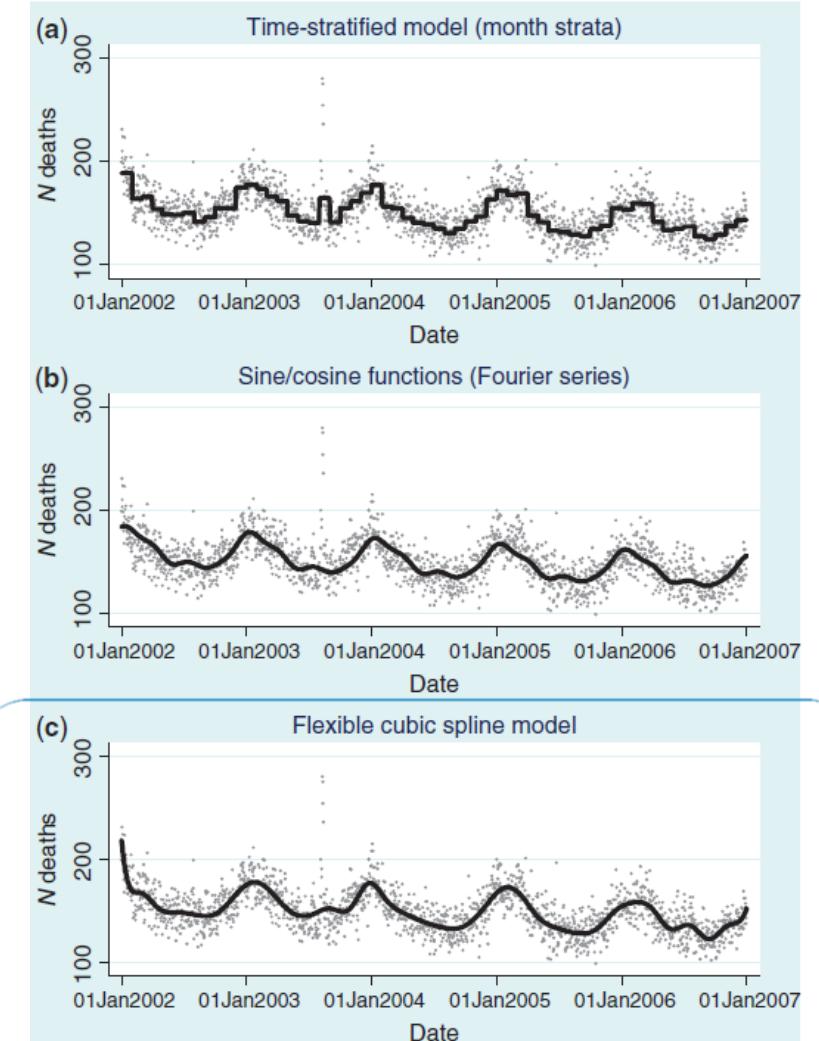


Three ways of modelling long-term pattern
(Figure 2 from Bhaskaran et al. 2013)

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

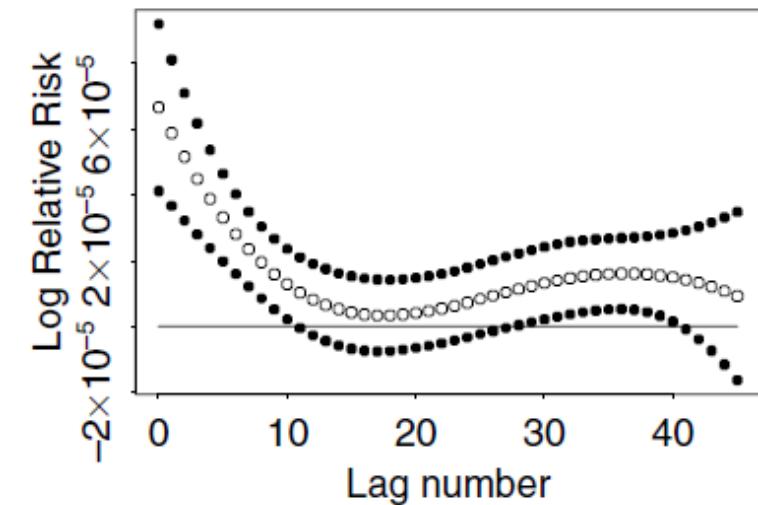


Three ways of modelling long-term pattern
(Figure 2 from Bhaskaran et al. 2013)

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 (linear or not)

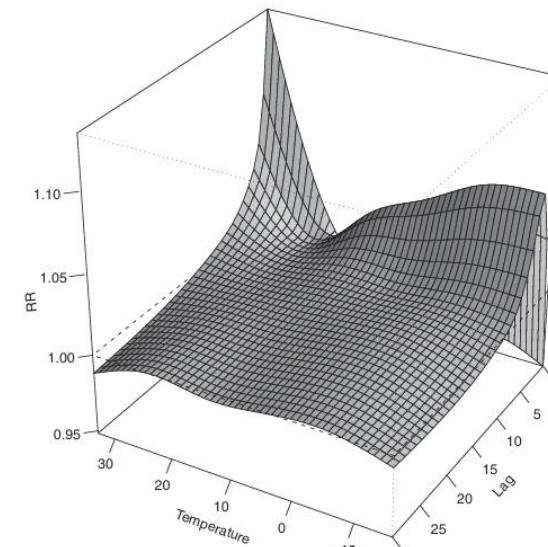
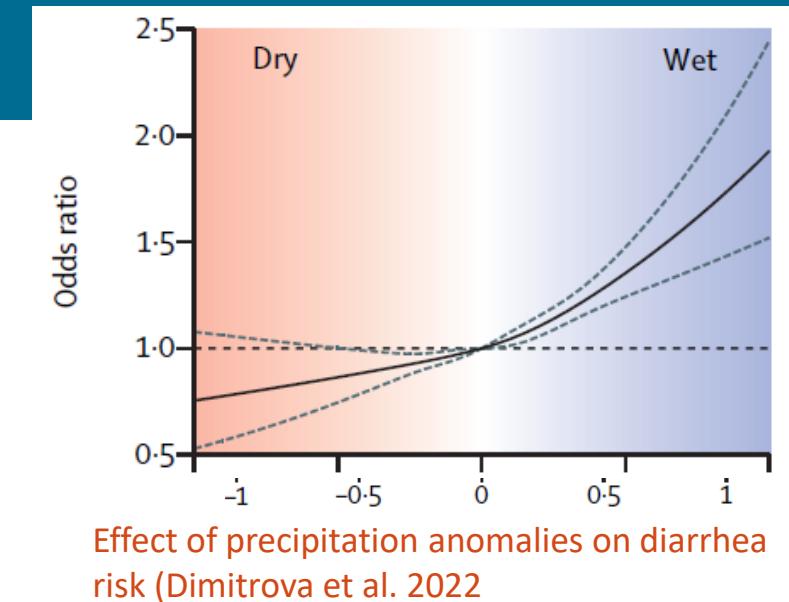


Association between TSP and mortality over lag structure with penalized spline function (Figure 3 from Zanobetti et al. 2000)



DISTRIBUTED LAG MODELS

- ❖ We need to consider both delayed and cumulative effects
 - ❖ **Delayed:** today's exposure may have an impact a few days later
 - ❖ **Cumulative:** consecutive exposure days are not independent and may synergistically increase risk
- ❖ Such lag effects can be linear or not
- ❖ The dose-response can also be non-linear (e.g. temperature)
- ❖ We can model both non-linear functions simultaneously
 - ❖ i.e. cross basis function
 - ❖ Be careful when using packages



Gasparrini A., Armstrong, B., Kenward M. G. Distributed lag non-linear models. Statistics in Medicine. 2010; 29(21):2224-2234.

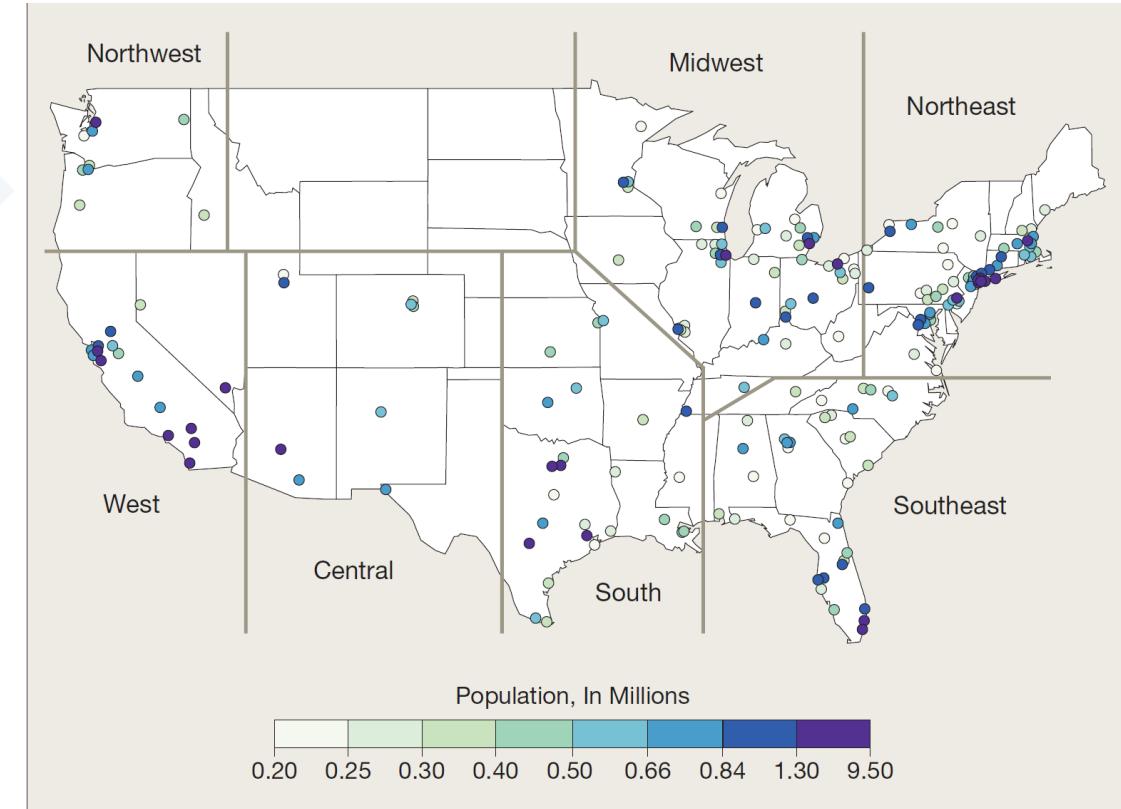
POOLING RESULTS ACROSS GEOGRAPHICAL UNITS

Assuming homogeneous concentration-response functions (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
 - ❖ Or stratifying by spatial units (case time series design)

What if heterogeneity exists?

- ❖ Effect modification: more next week



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	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	Latiun, 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

- thumb up icon flexibly applied to study the health impacts of either short- or long-term exposure to air pollution
- thumb up icon offers a feasible alternative when some other study designs are impractical or impossible to implement
- thumb up icon 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 $\text{PM}_{2.5}$ 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.

$$\begin{aligned} & \ln[E(\text{Mortality}_{c,m,y})] \\ &= \beta_1 \text{SmokePM}_{c,m,y} + \beta_2 \text{Temperature}_{c,m,y} + \beta_3 \text{Precipitation}_{c,m,y} + \text{offset}[\ln(\text{Population}_{c,y})] \\ &+ \alpha_c + \gamma_y + \varepsilon_{c,m,y}. \end{aligned}$$

- *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*
- $\ln(\text{Population}_{c,y})$: 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

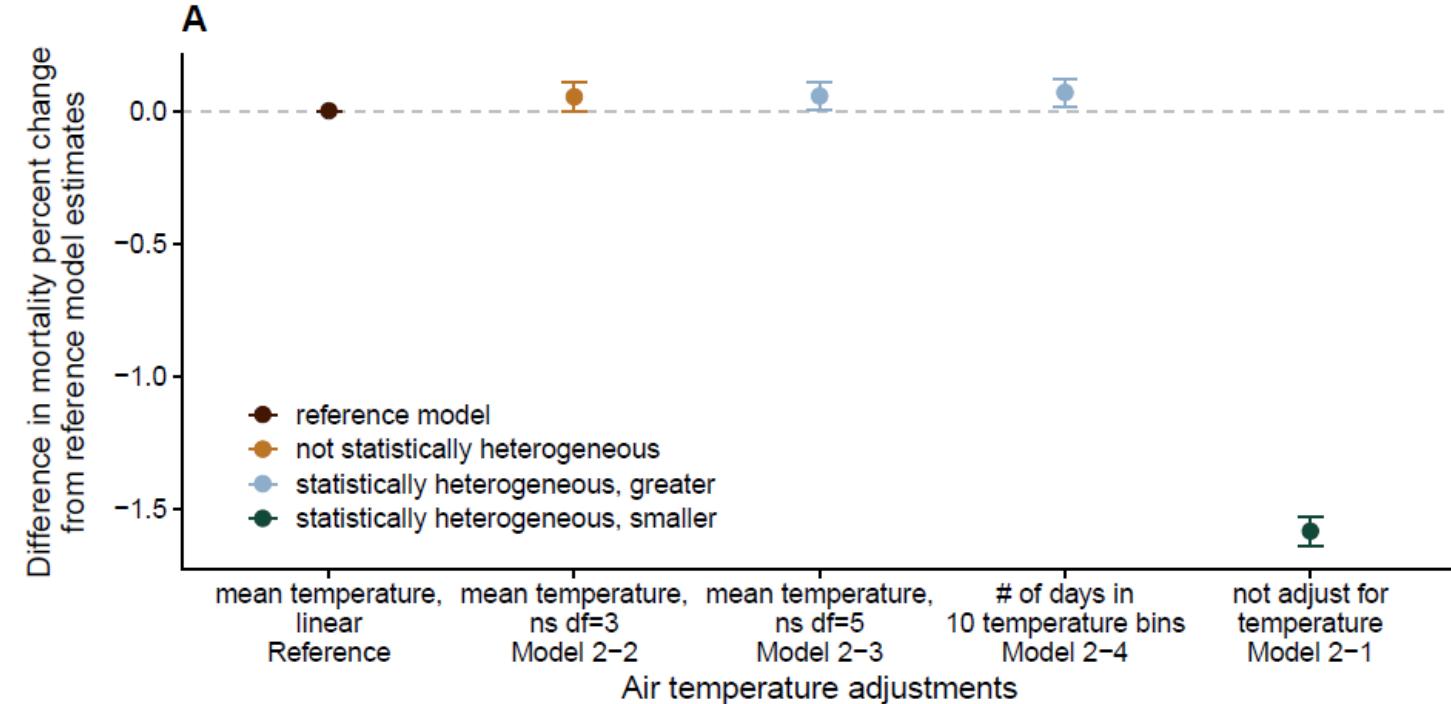
Each modified model is an alternative specification of the reference model, differing in only one aspect to isolate the effect of that particular change.

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

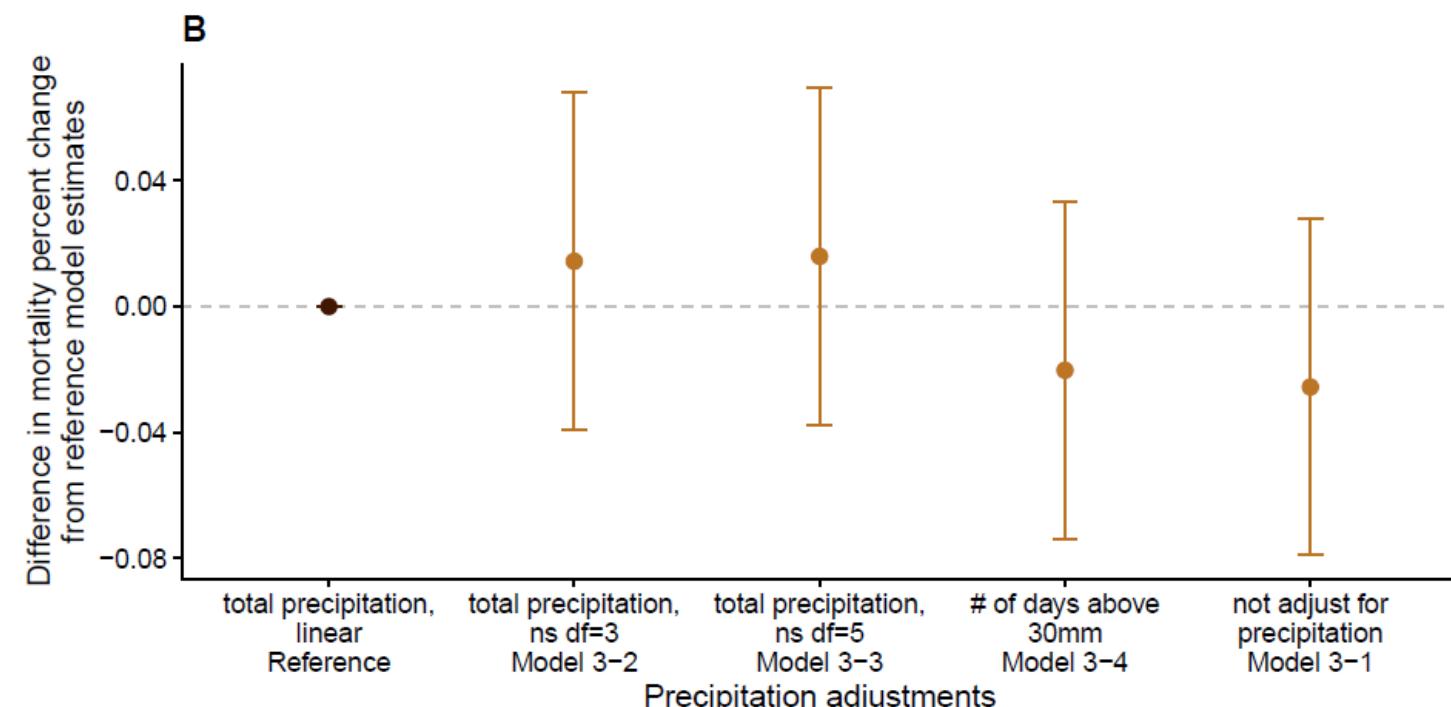
Note: SE, standard error; IID, independent and identically distributed; ns, natural cubic spline; df, degree of freedom.





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

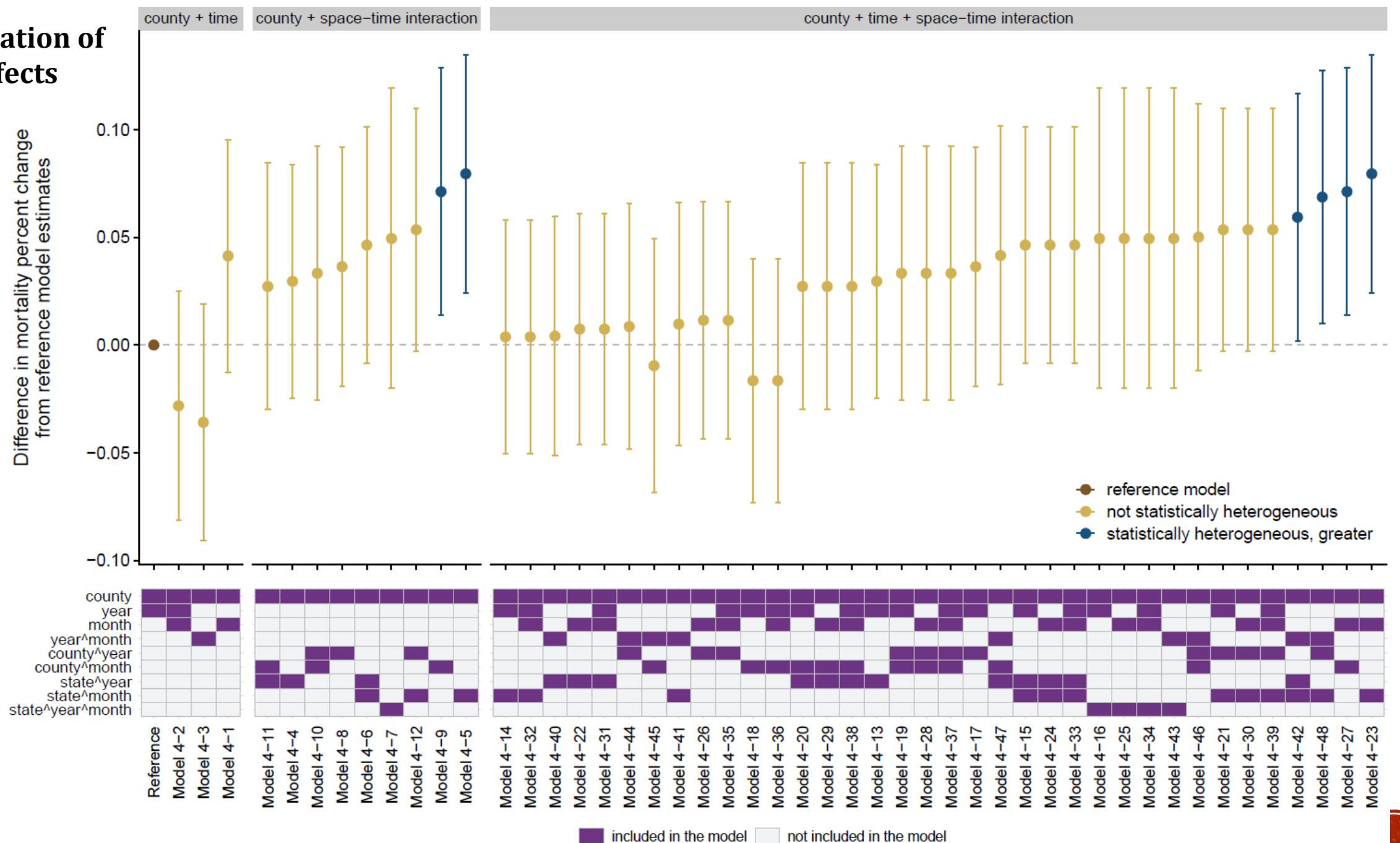


Figure 3. Differences in estimates between models with alternative combinations of fixed effects and the reference model.



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



Stage 2: Select a main model based on permutation test and model complexity

Step 5: Perform permutation tests for candidate models



Stage 3: Perform sensitivity analyses and consider uncertainty estimation methods

Step 9: Sensitivity analysis

Perform sensitivity analysis for the selected main model by reconsidering alternative specifications in Stage 1.

Step 10: Uncertainty estimation

Choose the appropriate method for SE calculation considering the data structure. For transparency, consider reporting results with different SE calculations.

statistically heterogenous

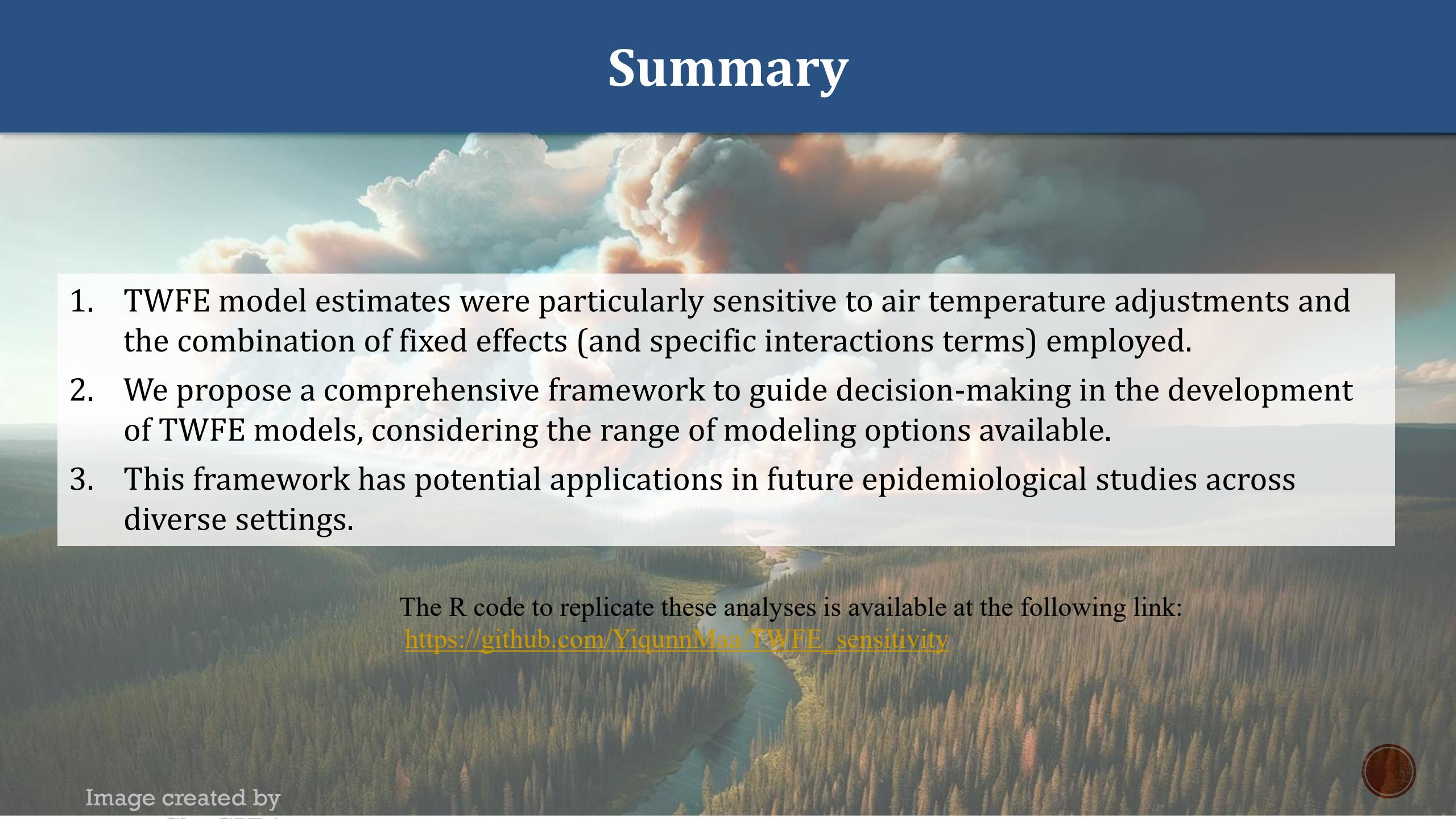
Exclude

not statistically heterogenous

Step 8: Select the simplest as the main model



Summary

- 
- The background of the slide features a wide-angle photograph of a natural landscape. In the foreground, there's a dense forest of tall evergreen trees. A winding river or stream flows through the center of the frame. The middle ground shows more forested hills. The sky above is filled with large, billowing clouds that have a mix of white, grey, and orange-tinted areas, suggesting either sunrise or sunset. The overall atmosphere is serene and expansive.
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



NATURAL EXPERIMENTS FOR CLIMATE EPIDEMIOLOGISTS



USING NATURAL EXPERIMENTS

- By capitalizing on natural experiments, Quasi-experimental methods (QEM) can be used as alternatives to experimental methods to provide causal estimates from observational studies.
- 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 et al. 1979)
- Different QEM, different configurations, different assumptions to draw causal inference
- To partially deal with both measured and unmeasured confounding



EXAMPLES OF NATURAL EXPERIMENTS NOT ONLY FOR POLICY EVALUATION

- Public policies:
 - Smoking ban
 - Legalization of marijuana
- Clinical Treatments, Vaccination
- Conditional Cash Transfers
- Natural Hazards
 - Earthquakes
 - Wildfires

2 main types of natural experiments

- Timing of the intervention
- Eligibility to a specific policy (e.g. age for vaccination) and other IV strategies



CONSIDERING CLIMATE HAZARDS AS NATURAL EXPERIMENTS

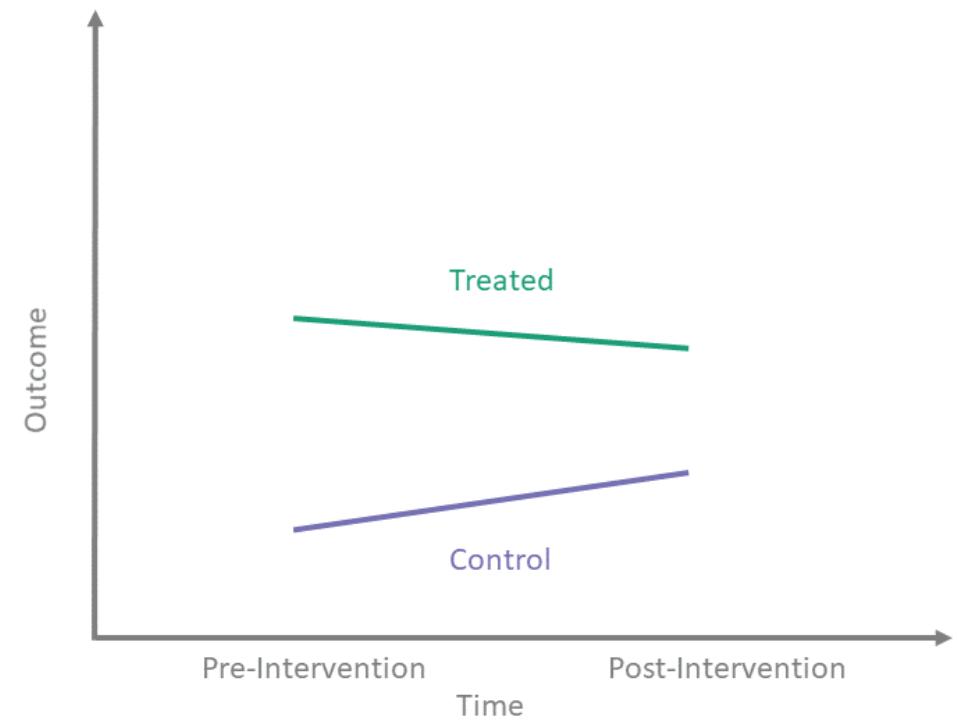
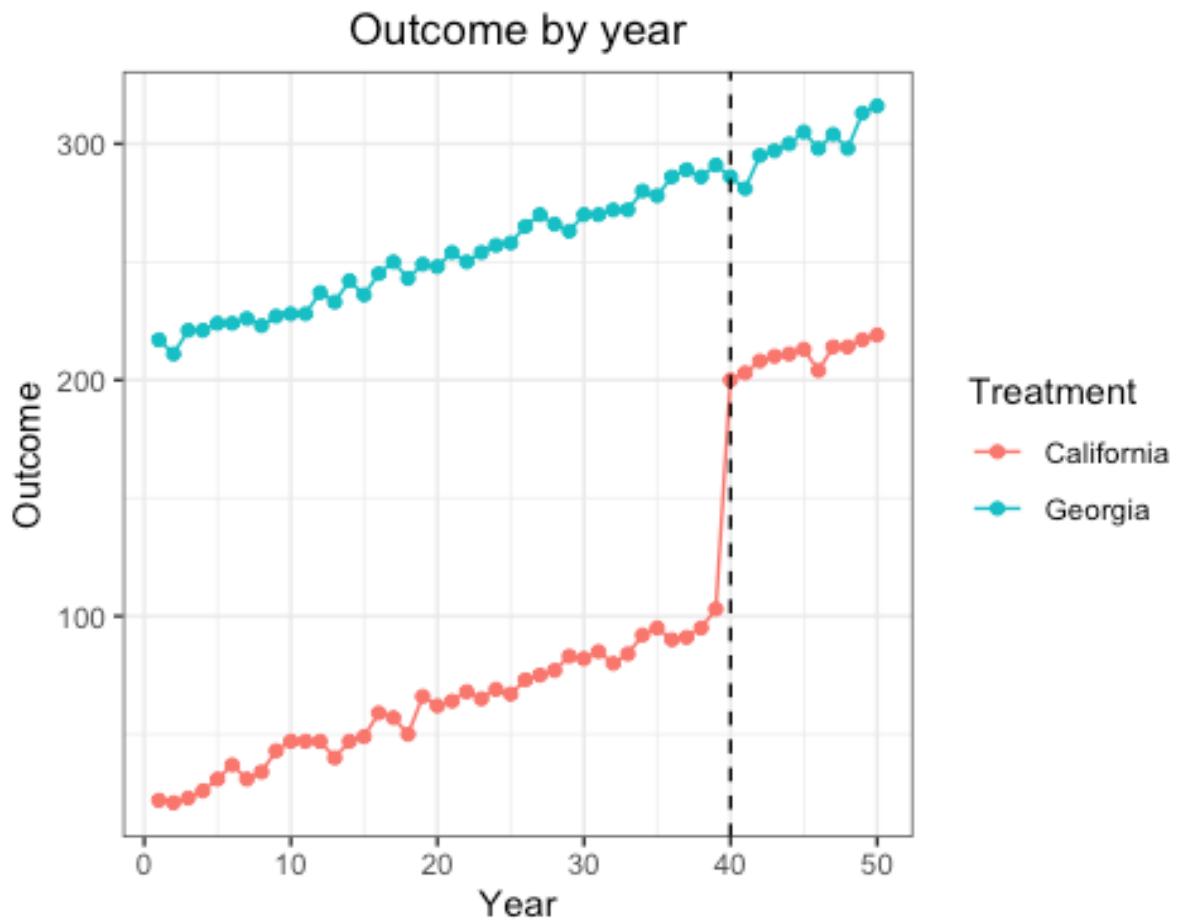
Using the Generalized Synthetic Control Method to Estimate the Impact of Extreme Weather Events on Population Health

Paige Sheridan,^{a,b} Sara McElroy,^{a,b} Joan Casey,^c and Tarik Benmarhnia^{a,d}

- If we can argue that the timing of a given extreme weather event is independent with the trend for the outcome of interest, we can capitalize on this as a natural experiment
- Multiple approaches exist to capitalize on such natural experiments
 - Difference-in-differences methods
 - Synthetic control methods
 - Interrupted time series analyses



The difference-in-differences idea



By Bret Zeldow and Laura Hatfield



DIFFERENCE-IN-DIFFERENCES

THE STANDARD APPROACH

- To estimate the effect of interest

$$(\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$$

$i = 0$ is control group, $i = 1$ is treatment.

$t = 0$ is pre-period, $t = 1$ is post-period.

- Simple Regression modeling approach (with 2 groups)

- $E[Y | a, t, did] = \beta_0 + \beta_1 a + \beta_2 t + \beta_3 did$

- Where:

- a represents the group with the policy
- t represents the period after the policy implementation
- DID is the interaction between a and t



HOW DOES IT WORK?

	T=0	T=1
A=0	β_0	$\beta_0 + \beta_1$
A=1	$B'0 + \beta_2$	$B'0 + \beta'1 + \beta'2 + \beta_3$

$$E[Y | a, t, did] = \beta_0 + \beta_1 t + \beta_2 a + \beta_3 [a*t]$$

Like a double or
2-way fixed effect

0 0
1 0
0 1
1 1

You can also
include more
complex time
trends

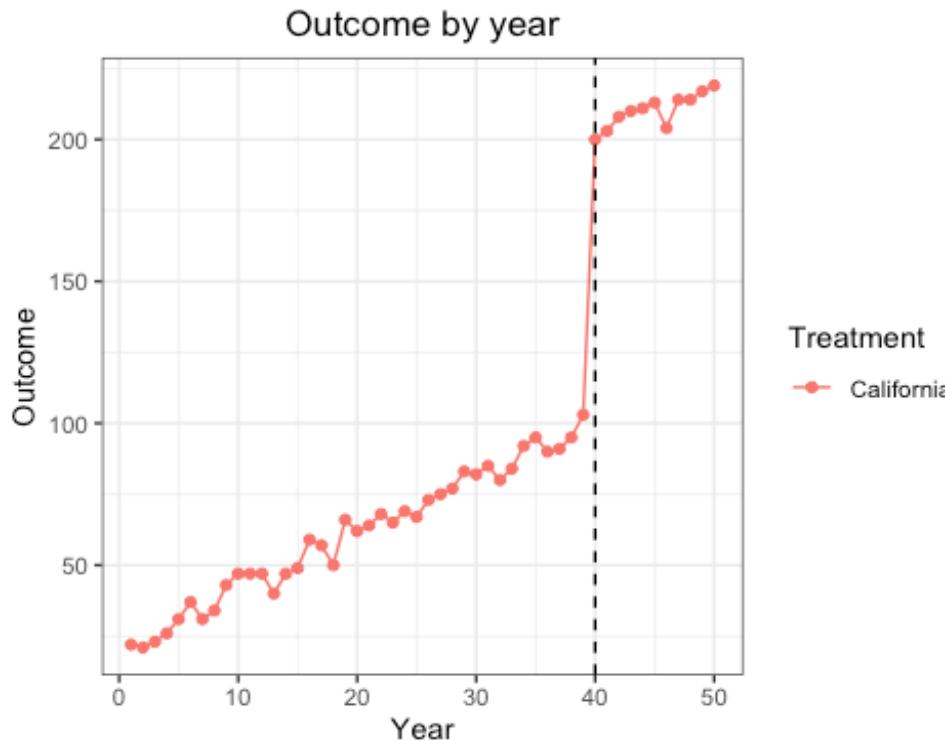


DID ASSUMPTIONS

- The key assumptions of the DiD analysis are:
 1. The trend in the control group represents a good approximation for the counterfactual trend of the treated group in the absence of the treatment.
 2. Common Shock Assumption
 3. No spillover



INTERRUPTED TIME SERIES: DID WITHOUT CONTROL GROUPS



Estimation

The traditional approach:

The following model can be used to estimate the effect of the policy on the outcome y

$$y = \alpha + \beta_1(\text{year}) + \beta_2(\text{post}) + \beta_3(x_i) + \beta_4(x_t) + \beta_5(x_{it}) + \beta_6(\text{year} \times \text{post}) + \epsilon$$

Can be also done through a 2-stage approach:

1. Building and optimizing a predictive model for Y in the pre-treatment period
2. Predicting Y in the post-treatment period using the model developed in stage 1 and compare with observed outcomes

Various approaches can be used for stage 1 (ARIMA, random forest, and other ML algorithms)



TWO-STAGE TIME SERIES ANALYSIS COUPLED WITH MACHINE LEARNING: EVALUATING THE HEALTH EFFECTS OF THE 2018 WILDFIRE SMOKE EVENT IN SAN FRANCISCO COUNTY AS A CASE STUDY

- We compared multiple algorithms in a 2-stage ITS approach
 - Autoregressive Integrated Moving Average (ARIMA)
 - NNETAR (Neural Network)
 - Prophet-XGBoost

Table 1. Performance metrics of the ARIMA, NNETAR, and Prophet-XGBoost model

	Training (2009-01-01 to 2016-11-07)			Testing (2016-11-08 to 2018-11-07)		
	ARIMA	NNETAR	Prophet-XGBoost	ARIMA	NNETAR	Prophet-XGBoost
R ²	0.71	0.83	0.83	0.65	0.58	0.71
MAE	7.08	5.52	5.33	8.67	9.36	8.11
RMSE	9.19	7.12	7.02	11.84	13.00	10.85
MAPE	0.14	0.11	0.11	0.16	0.16	0.15
SMAPE	0.13	0.11	0.10	0.16	0.17	0.15

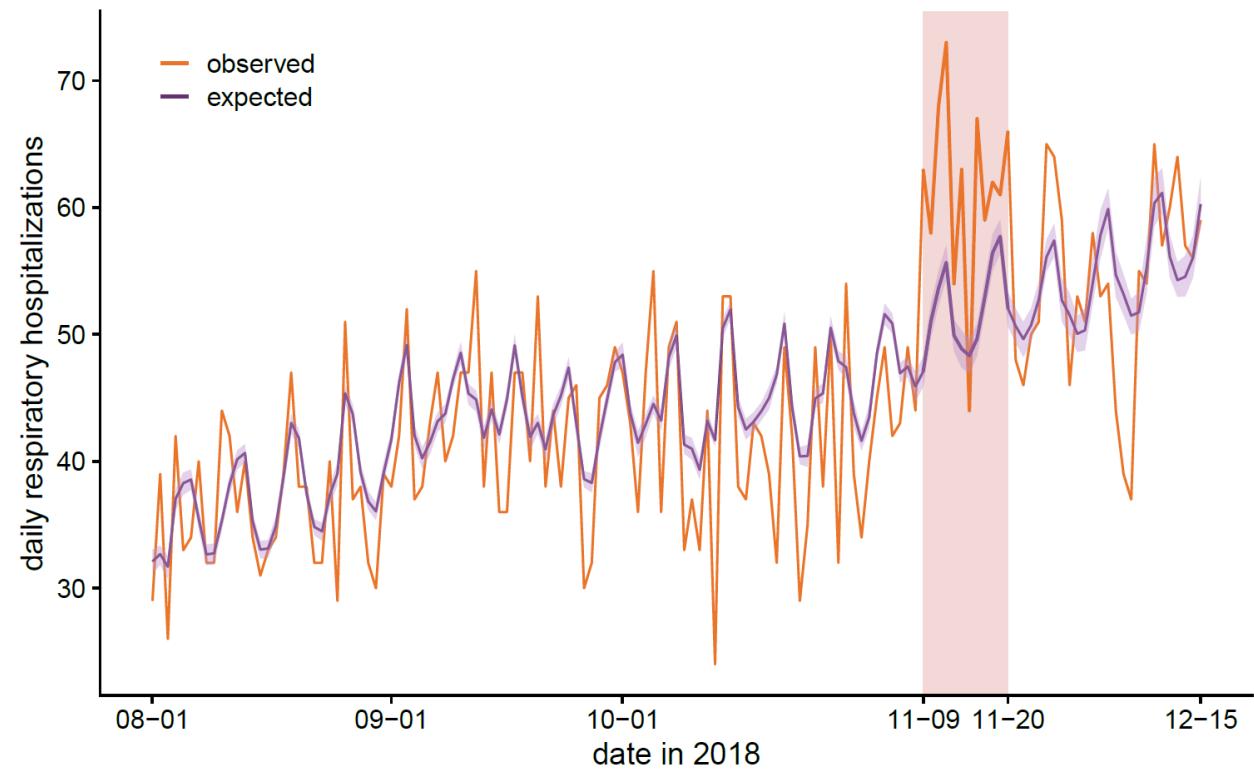
R²: coefficient of determination

MAE: mean absolute error

RMSE: root mean square error

MAPE: mean absolute percentage error

SMAPE: symmetric mean absolute percentage error



Acknowledgements

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

Sigal Maya and the Workshop Organizing Committee

Sheri Weiser, Arianne Teherani and the P20 team



THANK YOU

YIM022@UCSD.EDU
TBENMARHNIA@UCSD.EDU

