QUASI-EXPERIMENTAL METHODS FOR CLIMATE EPIDEMIOLOGY PART III

21 March 2025

Tarik Benmarhnia & Yiqun Ma

University of California, San Diego

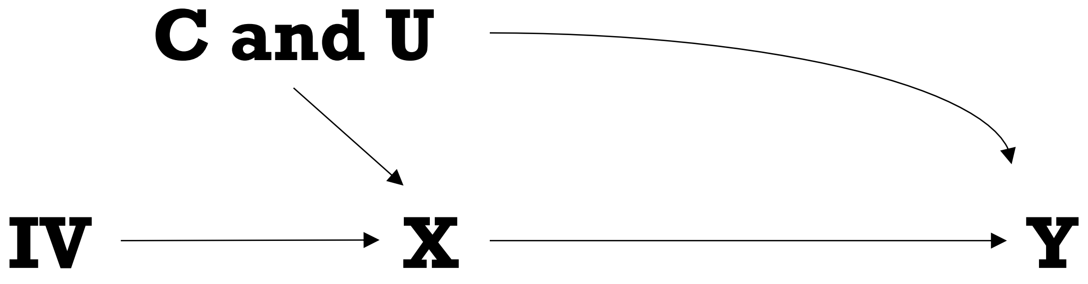
Drexel Climate Change and Urban Health Research Center Workshop

OUTLINE

- Internal and external validity
  - The importance of effect (measure) modification
  - Transportability/Generalizability
- Natural experiments based on eligibility
  - IV methods
  - Regression Discontinuity
- Natural experiments based on timing, advanced topics:
  - staggered interventions and multi-stage models for heterogeneous effects

THE IDEA BEHIND RANDOMIZATION

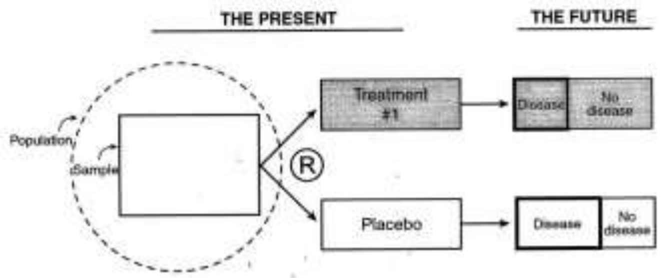
- How do RCT work …
- Randomization can be analyzed as an "**Instrumental Variable**"
- The overall aim is to deal with both measured and unmeasured confounding



- **The problem:** in many cases RCT are not feasible due costs, scale or ethical objections
- Or because it is simply too late ..

DIFFERENT TYPES OF RCTS

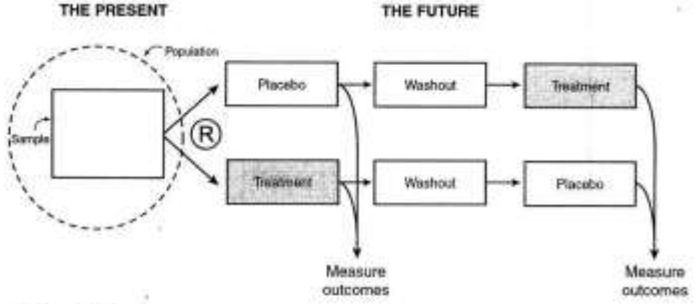
- Two-arm, parallel design
- Planned cross-over design
- Factorial design
- Cluster randomized trials
- Stepped Wedge Designs



■ **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.

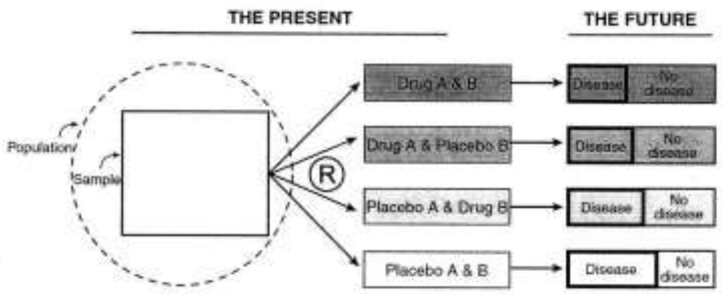
Two-arm, parallel design



■ **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
  - **A side note on transportability/generalizability**

Contents lists available at ScienceDirect

Social Science & Medicine

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

ELSEVIER

Understanding and misunderstanding randomized controlled trials

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

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

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.

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 <u>different effect measures</u> (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 <u>measure</u> 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
   - <u>Additive, but not multiplicative</u>, 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 β_1 = ln (RRstrata$_1$); β_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)

An overview of modern machine learning methods for effect measure modification analyses in high-dimensional settings

Michael Cheung, Anna Dimitrova, Tarik Benmarhnia

Scripps Institution of Oceanography, University of California, San Diego, CA, USA

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?

EXTERNAL VALIDITY AND TARGET POPULATION

- Internal validity refers to the validity of a causal effect estimate in the study sample → BIAS
- "No causal effect is fully specified unless we define a target population for that causal effect" (Maldonado & Greenland 2002)
- Target population:
  - The population in whom we ultimately want causal knowledge
  - Which population do we target when we make recommendations?
- External validity refers to the similarity between the study population (sample) and the target population
- In this context, we distinguish 2 types of external validity:
  - Generalizability
  - Transportability

Maldonado, G., & Greenland, S. (2002). Estimating causal effects. International journal of epidemiology, 31(2), 422-429.

THE IMPORTANCE OF EFFECT MODIFIERS

- The main reason for which an effect estimate in a given study may not be generalizable (to the target population) or transportable is because of a **differential distribution of effect modifiers**
- And effect (measure) modifiers are not uncommon (on a given scale)

CASE STUDY, WHY EMM COMPOSITION MATTERS

A RCT ABOUT A TREATMENT *T* ON THE 5-YEARS RISK ON STROKE
GENDER IS AN EFFECT MODIFIER (SEE P. 210-211)

Table 9.1 DATA CONSISTENT WITH FIGURE 9.4

Men	Stroke	No stroke	Total	Risk	RD
Treated	75	1,425	1,500	5%	−15%
Untreated	300	1,200	1,500	20%	
Women	**Stroke**	**No stroke**	**Total**	**Risk**	**RD**
Treated	25	475	500	5%	−5%
Untreated	50	450	500	10%	
All	**Stroke**	**No stroke**	**Total**	**Risk**	**RD**
Treated	100	1,900	2,000	5%	−12.5%
Untreated	350	1,650	2,000	17.5%	

- We can calculate the weighted average RD across gender strata for table 1
- RD = (15% x [3000/4000]) + (5% x [1000/4000]) = **12.5%**

Table 9.2 DATA CONSISTENT WITH FIGURE 9.5

Men	Stroke	No stroke	Total	Risk	RD
Treated	75	1,425	1,500	5%	−15%
Untreated	300	1,200	1,500	20%	
Women	**Stroke**	**No stroke**	**Total**	**Risk**	**RD**
Treated	75	1,425	1,500	5%	−5%
Untreated	150	1,350	1,500	10%	
All	**Stroke**	**No stroke**	**Total**	**Risk**	**RD**
Treated	150	2,850	3,000	5%	−10%
Untreated	450	2,550	3,000	15%	

- We can calculate the weighted average RD across gender strata for table 1
- RD = (15% x [3000/6000]) + (5% x [3000/4000]) = **10%**

REVISITING IDENTIFICATION ASSUMPTIONS FOR EXTERNAL VALIDITY

- External Exchangeability:
  - No differences in distributions of effect measure modifiers between the study sample and the target population [p.208]
- External Positivity:
  - For example, if we have no Hispanic women over 65 in the study sample and the target population includes such population subgroup, we have an issue of external positivity
  - Such issue (as for internal validity positivity) can be addressed by model extrapolation but with some strong (and sometimes unverifiable) assumptions.
- External validity implications of the consistency violation:
  - If the exposure of interest is not consistent, the inference may not be generalizable to the target population
  - Ex: Hernán, M. A., & Taubman, S. L. (2008). *Does obesity shorten life? The importance of well-defined interventions to answer causal questions. International journal of obesity, 32(3), S8-S14*.
  - Different interventions can reduce BMI: Physical activity, Diet, Surgery
  - If there are EMMs and the distribution of EMMs differs between the study sample and the target population, results will not be generalizable.

REVISITING THE HIERARCHY OF STUDY DESIGNS

- While ideal RCT (that do not exist in practice) aims at reaching exchangeability, it remains an internal validity issue
- As we should always have a target population in mind, the **<u>gold standard for causal inference should be a doubly randomized trial</u>**
  - Random selection from the target population + random treatment assignment

AN OVERVIEW OF EXISTING APPROACHES

- Many methods exist to estimate population intervention effects
  - You already know how to calculate the population attributable fraction
  - More recent approaches have been proposed and offer more flexibility
    - To have different scenarios (beyond moving all exposed to unexposed only)
    - To deal with time-varying exposures and time varying confounding
    - To consider effect measure modification
  - For example, parametric g-computation

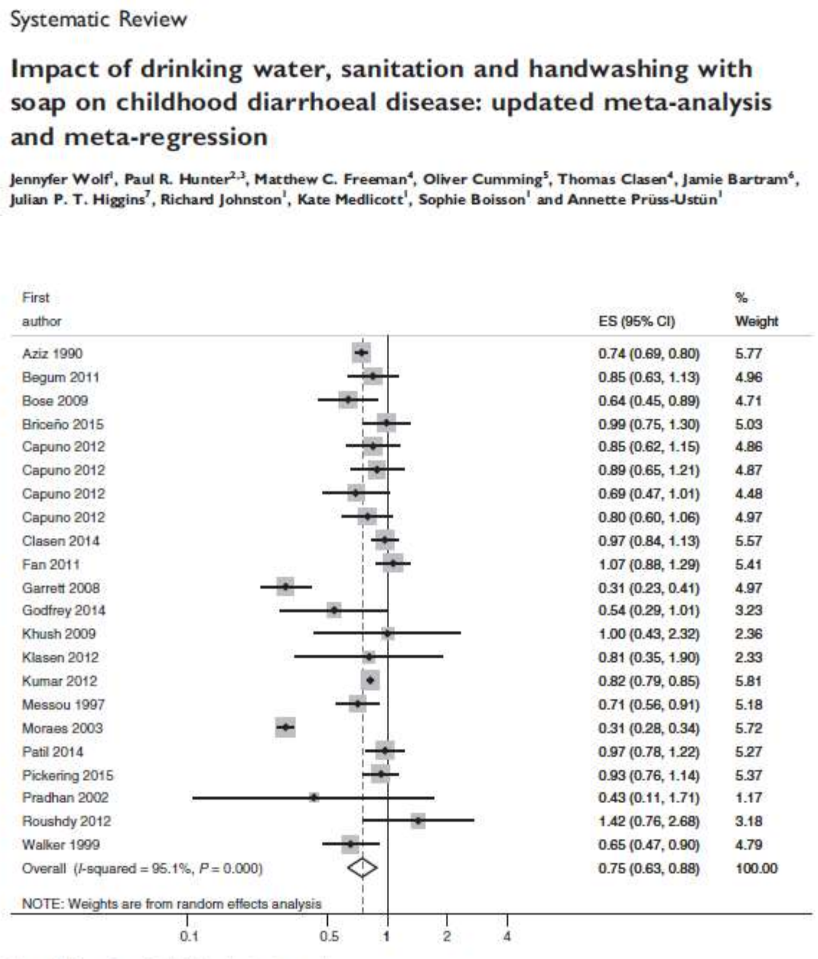
$$PAF = \frac{Pe \times (RR - 1)}{[Pe \times (RR - 1)] + 1}$$

Where:
Pe is total population exposed in %
RR the risk ratio which measures the association between the exposure and the disease of interest.

- For generalizability
  - Different approaches, based on inverse probability weighting, exist to generalize results from a given study to the target population
  - You probably already know survey weights = IPCW
  - Recently applied in the context of such issues (see EHR example later)
- For transportability
  - Traditional approaches include health impact assessment
  - More recent approaches have been developed for transportability
    - Inverse Odds of Sampling Weights
    - Substitution estimator, G-computation

Modeling the Impact of Population Intervention Strategies on Reducing Health Disparities: Water, Sanitation, and Hygiene Interventions and Childhood Diarrheal Disease in Peru

Rudy Patrick,[1]* Sara McElroy,[2,3] Lara Schwarz,[2,3] Georgia Kayser,[2] and Tarik Benmarhnia[2,3]



Systematic Review

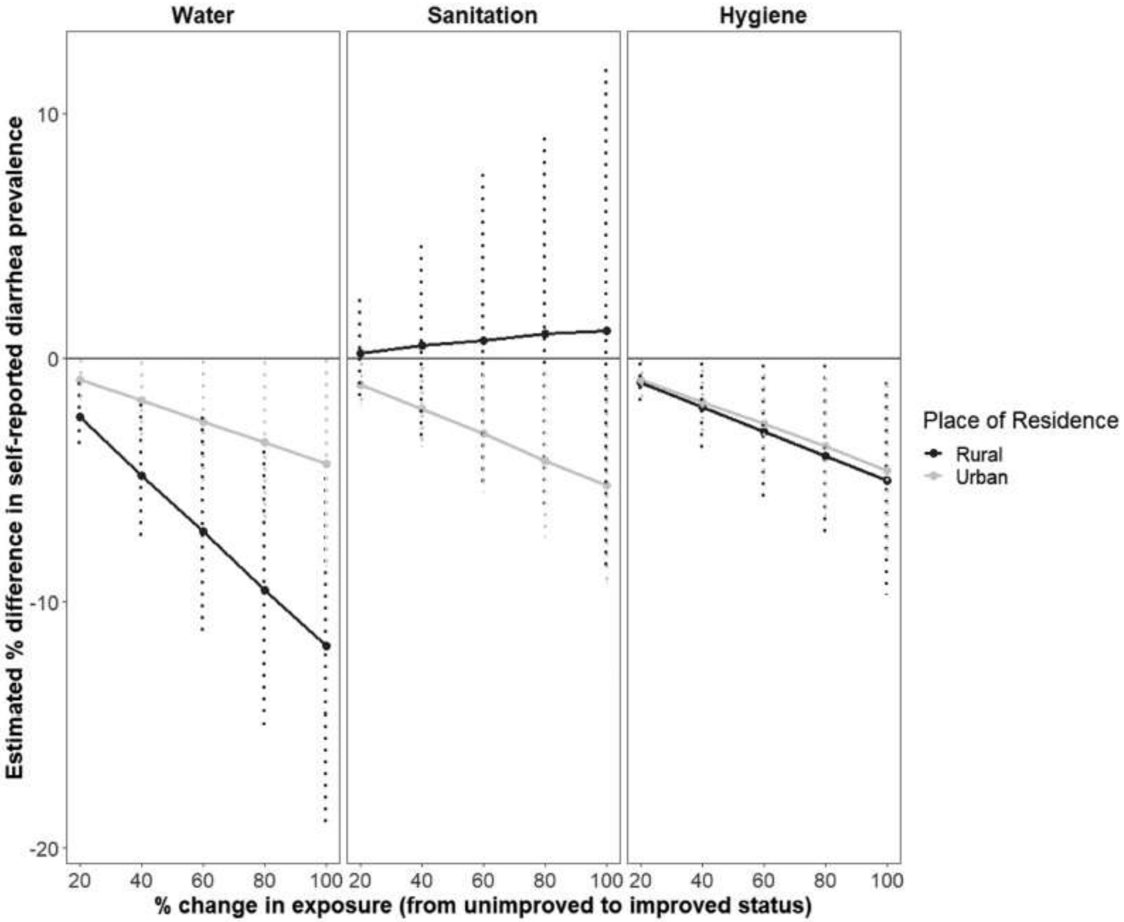
Impact of drinking water, sanitation and handwashing with soap on childhood diarrhoeal disease: updated meta-analysis and meta-regression

Jennyfer Wolf[1], Paul R. Hunter[2,3], Matthew C. Freeman[4], Oliver Cumming[5], Thomas Clasen[4], Jamie Bartram[6], Julian P. T. Higgins[7], Richard Johnston[1], Kate Medlicott[1], Sophie Boisson[1] and Annette Prüss-Ustün[1]

First author	ES (95% CI)	% Weight
Aziz 1990	0.74 (0.69, 0.80)	5.77
Begum 2011	0.85 (0.63, 1.13)	4.96
Bose 2009	0.64 (0.45, 0.89)	4.71
Briceño 2015	0.99 (0.75, 1.30)	5.03
Capuno 2012	0.85 (0.62, 1.15)	4.86
Capuno 2012	0.89 (0.65, 1.21)	4.87
Capuno 2012	0.69 (0.47, 1.01)	4.48
Capuno 2012	0.80 (0.60, 1.06)	4.97
Clasen 2014	0.97 (0.84, 1.13)	5.57
Fan 2011	1.07 (0.88, 1.29)	5.41
Garrett 2008	0.31 (0.23, 0.41)	4.97
Godfrey 2014	0.54 (0.29, 1.01)	3.23
Khush 2009	1.00 (0.43, 2.32)	2.36
Klasen 2012	0.81 (0.35, 1.90)	2.33
Kumar 2012	0.82 (0.79, 0.85)	5.81
Messou 1997	0.71 (0.56, 0.91)	5.18
Moraes 2003	0.31 (0.28, 0.34)	5.72
Patil 2014	0.97 (0.78, 1.22)	5.27
Pickering 2015	0.93 (0.76, 1.14)	5.37
Pradhan 2002	0.43 (0.11, 1.71)	1.17
Roushdy 2012	1.42 (0.76, 2.68)	3.18
Walker 1999	0.65 (0.47, 0.90)	4.79
Overall (*I*-squared = 95.1%, *P* = 0.000)	0.75 (0.63, 0.88)	100.00

NOTE: Weights are from random effects analysis

Figure 5 Forest plot of included sanitation interventions.



NATURAL EXPERIMENTS BASED ON ELIGIBILITY

GENERAL IDEA

$$
\begin{tikzcd}
& \textbf{C and U} \arrow[d] \arrow[dr, bend left] & \\
\textbf{IV} \arrow[r] & \textbf{X} \arrow[r] & \textbf{Y}
\end{tikzcd}
$$

- The overall aim is to deal with measured and unmeasured confounding in observational studies by using an instrumental variable
- The idea behind instrumental variables (IVs) is that there is a perfect "exogenous" variable that is correlated with the exposure variable of interest (X), but which has no effect on the outcome variable (Y) other than through X.
- Reminder: Randomization can be seen as an IV
- Long tradition in econometrics (Wright, 1928; Golderberger, 1972)
- Recently formalized using the counterfactual (or potential outcomes) framework (Angrist, 1994; Robbins, 1994)
- More Recently used in public health (Hernan 2006)

SOME IV EXAMPLES

- Differential distance
  - (home to high-tech hospital) – (home to nearest hospital)
  - Caniglia, Ellen C., et al. "Methodological challenges when studying distance to care as an exposure in health research." American journal of epidemiology 188.9 (2019): 1674-1681.
- Law changes
- **Eligibility for social programs (Income, age, etc..)**
- Genes (Mendelian Randomization)
  - Lawlor, Debbie A., et al. "Mendelian randomization: using genes as instruments for making causal inferences in epidemiology." *Statistics in medicine* 27.8 (2008): 1133-1163.
- Amount of rainfall received
- **Clinical guidelines (hypertension, etc.)**
- Plant closures
- Geographic location of railway lines
- **<u>Time</u>**

IV ESTIMATION
THE 2-STAGE LEAST SQUARES (2SLS) PROCEDURE

- Suppose a continuous outcome Y, an continuous exposure of interest X, an IV Z, and Cn are confounders
- Fitting consecutively two linear models:
- **Stage 1**: fit a linear regression of X on Z and C_1 and compute the predicted values
  - $E[X| Z, C1] = \beta_0 + \beta_1 z + \beta_2 c_1$
- **Stage 2**: fit a linear regression of Y on $\hat{E}$ $(X|Z,C_1)$ and C_2
  - $E[Y| \hat{E}, C_2] = \beta_0 + \beta_1 \hat{E} + \beta_2 c_2$

ASSUMPTIONS

- Can be summarized as follows (Jackson & Swanson 2015):

  (1) The IV has to be associated with the exposure/treatment;

  (2) The IV shares no common causes with the outcome

  (3) Any effect the IV has on the outcome is fully mediated by treatment (**exclusion restriction**).

- These 3 assumptions need to be met in order to obtain a valid IV estimator

REVISITING SOME IV EXAMPLES

- Differential distance
  - (home to high-tech hospital) – (home to nearest hospital)
  - Caniglia, Ellen C., et al. "Methodological challenges when studying distance to care as an exposure in health research." American journal of epidemiology 188.9 (2019): 1674-1681.
- Geographic location of railway lines
- Meteorological or other exogenous environmental events (e.g. thermal inversions, changes in the local wind direction, wildfire smoke)

Original article

Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk

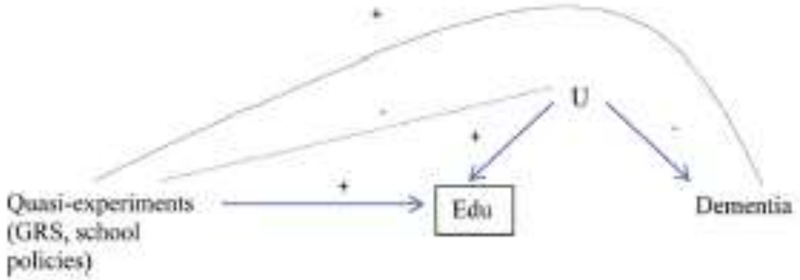
Thu T. Nguyen ScD, MSPH [a,*], Eric J. Tchetgen Tchetgen PhD [b,c], Ichiro Kawachi MD, PhD [d], Stephen E. Gilman ScD [c,d,e], Stefan Walter PhD [a], Sze Y. Liu PhD [f], Jennifer J. Manly PhD [g], M. Maria Glvmour ScD [a]

A B S T R A C T

Purpose: Education is an established correlate of cognitive status in older adulthood, but whether expanding educational opportunities would improve cognitive functioning remains unclear given limitations of prior studies for causal inference. Therefore, we conducted instrumental variable (IV) analyses of the association between education and dementia risk, using for the first time in this area, genetic variants as instruments as well as state-level school policies.
Methods: IV analyses in the Health and Retirement Study cohort (1998–2010) used two sets of instruments: (1) a genetic risk score constructed from three single-nucleotide polymorphisms (SNPs; $n = 7981$); and (2) compulsory schooling laws (CSLs) and state school characteristics (term length, student teacher ratios, and expenditures; $n = 10{,}955$).
Results: Using the genetic risk score as an IV, there was a 1.1% reduction in dementia risk per year of schooling (95% confidence interval, −2.4 to 0.02). Leveraging compulsory schooling laws and state school characteristics as IVs, there was a substantially larger protective effect (−9.5%; 95% confidence interval, −14.8 to −4.2). Analyses evaluating the plausibility of the IV assumptions indicated estimates derived from analyses relying on CSLs provide the best estimates of the causal effect of education.
Conclusions: IV analyses suggest education is protective against risk of dementia in older adulthood.

$$\begin{aligned}1^{st}\text{stage}: \text{Predicted}_{education} = {} & \beta_0 + \beta_1 \text{CSL} + \beta_2 \text{CSLw} \\ & + \beta_3 \text{Term\_length} \\ & + \beta_4 \text{Student-teacher\_ratio} \\ & + \beta_5 \text{Per-pupil\_expenditures} \\ & + \beta_6 \text{Student-teacher\_ratio} + \beta_7 C\end{aligned} \quad (1)$$

$$2^{nd}\ \text{stage}: \text{Dementia risk} = \eta_0 + \eta_1 \text{Predicted}_{education} + \eta_2 C + e \quad (2)$$



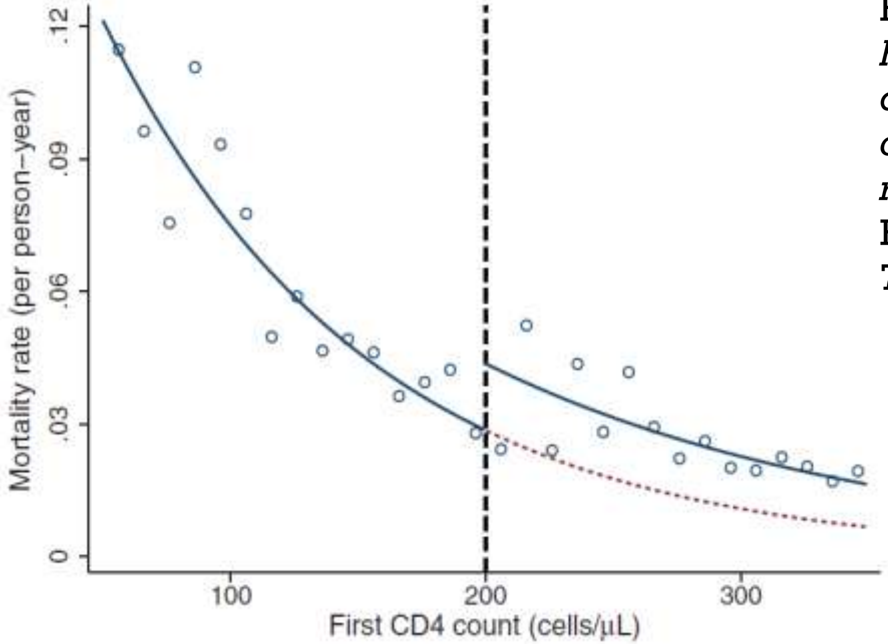
eFigure 1. Direct acyclic graph of relationship between instruments, years of schooling, and dementia

Table 4
CA and IV estimates for the effect of years of schooling on dementia using CSLs and school characteristics and education genetic risk score (GRS) as instruments

Instrument	IV estimate				Covariate adjustment estimate			
	N	β_{IV}	95% CI	*P*	β_{CA}	95% CI	*P*	Wu-Hausman, *P*[*]
Years of Schooling[†]	10,955	−0.095	(−0.148, −0.042)	<.001	−0.021	(−0.023, −0.019)	<.001	.006
Edu GRS[‡]	7981	−0.011	(−0.024, 0.002)	.11	−0.005	(−0.006, −0.004)	<.001	.39

THE REGRESSION DISCONTINUITY DESIGN

- A specific type of instrumental variable
- Whenever a decision rule assigns treatment to ‘patients’ who score higher (or lower) than a particular cutoff value on a continuously measured variable
- Some assumptions and validity conditions



Bor, J., et al. (2014). *Regression discontinuity designs in epidemiology: causal inference without randomized trials*. Epidemiology, 25(5), 729-737.

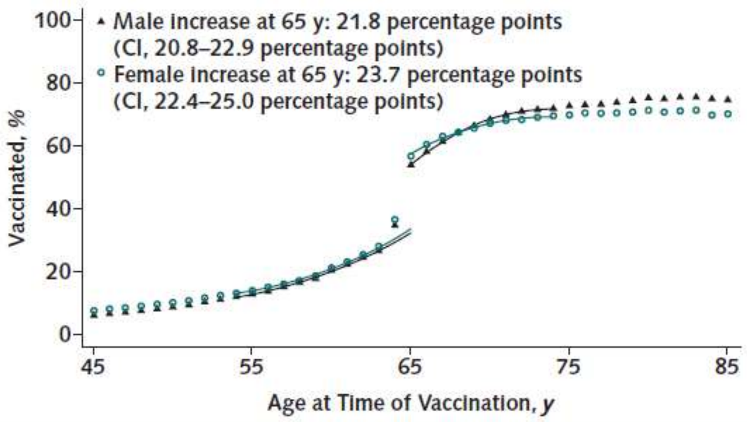
VACCINATION AS AN EXAMPLE

The Effect of Influenza Vaccination for the Elderly on Hospitalization and Mortality

An Observational Study With a Regression Discontinuity Design

Michael L. Anderson, PhD; Carlos Dobkin, PhD; and Devon Gorry, PhD

Figure 1. Age profile of vaccination status.



Age is calculated at the time of vaccination. Vaccinations are given in September and October, so a substantial fraction of persons aged 64 y will turn 65 during the influenza season. For this reason, the regressions were fitted dropping the cell at 64 y. (Data from Royal College of General Practitioners 2003-2012.)

Evaluating the Effectiveness of Vaccines Using a Regression Discontinuity Design

Nicole E. Basta* and M. Elizabeth Halloran

MANY EXAMPLES

Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy

Yuyu Chen[a,1], Avraham Ebenstein[b,1], Michael Greenstone[c,d,1,2], and Hongbin Li[a,1]

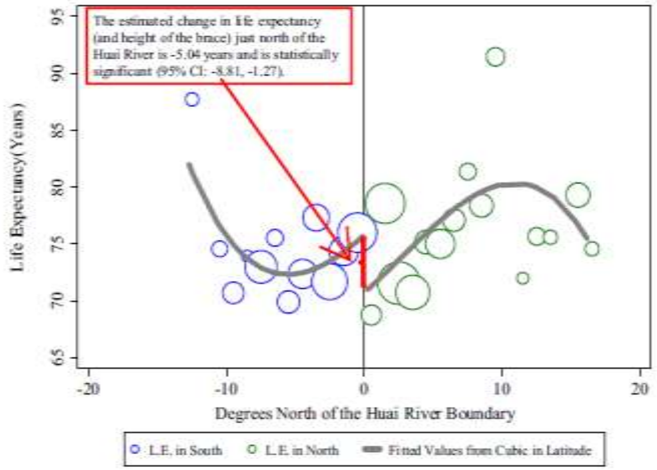


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

ELSEVIER
Contents lists available at SciVerse ScienceDirect
Social Science & Medicine
journal homepage: www.elsevier.com/locate/socscimed
SOCIAL SCIENCE & MEDICINE

The impact of drinking on psychological well-being: Evidence from minimum drinking age laws in the United States

Ceren Ertan Yörük[a,*], Barış K. Yörük[b]

Government health insurance for people below poverty line in India: quasi-experimental evaluation of insurance and health outcomes

thebmj

Neeraj Sood *associate professor*[1,2,3], Eran Bendavid *assistant professor*[4,5], Arnab Mukherji *associate professor*[6], Zachary Wagner *PhD student*[7], Somil Nagpal *senior health specialist*[8], Patrick Mullen *senior health specialist*[8]

REVIEW OF THE LITERATURE (IN 2015)

Table 1. PubMed articles with health outcomes using regression discontinuity designs

Authors	Year	Journal	Study topic
Albouy and Lequien [23]	2009	Journal Health Economics	Effect of education on mortality
Almond et al. [15]	2010	Quarterly Journal of Economics	Returns to treatment of low-birth-weight newborns
Andalón [24]	2011	Health Economics	Effect of Oportunidades on obesity
Anderson et al. [25]	2011	Journal of Health Economics	Effect of schooling on children's BMI
Arcand and Wouabe [26]	2010	Health Economics	Effect of teacher training on HIV prevention
Banks and Mazzonna [27]	2012	Economics Journal	Effect of education on old-age cognitive ability
Behrman [28]	2014	Social Science and Medicine	Effect of primary schooling on HIV status
Bor et al. [1]	2014	Epidemiology	Effect of early vs. deferred HIV treatment on mortality
Callaghan et al. [29]	2014	Drug and Alcohol Dependence	Effect of legal drinking age on mortality
Callaghan et al. [30]	2013	American Journal of Public Health	Effect of legal drinking age on alcohol-related morbidity
Callaghan et al. [31]	2013	Addiction	Effect of legal drinking age on inpatient morbidity
Carpenter and Dobkin [16]	2009	AEJ: Applied Economics	Effect of alcohol consumption on mortality
Carpenter and Dobkin [32]	2011	Journal of Economic Perspectives	Minimum legal drinking age and public health
Chen et al. [33]	2013	PNAS	Effect of air pollution on mortality
Conover and Scrimgeour [34]	2013	Journal of Health Economics	Health effects of minimum legal drinking age
De La Mata [35]	2012	Health Economics	Effect of Medicaid eligibility on coverage, utilization, and health
Deza [36]	2014	Health Economics	Effect of alcohol use on drug consumption
Flam-Zalcman et al. [37]	2012	Intl J Psych Research	Effect of criterion-based increase in alcohol treatment
Fletcher [38]	2014	Biodemography and Social Biology	Effect of genetics on stress response
Glance et al. [39]	2014	JAMA Surgery	Effect of hospital report cards on mortality
Gormley et al. [40]	2005	Developmental Psychology	Effect of universal pre-kindergarten on cognitive development
Huang and Zhou [41]	2013	Social Science and Medicine	Effect of education of cognition
Jensen and Wust [42]	2014	Journal of Health Economics	Effect of Caesarean section on maternal and child health
McFarlane et al. [43]	2014	Schizophrenia Bulletin	Effect of treatment program on psychosis onset
Miller et al. [44]	2013	AEJ: Applied Economics	Effect of insurance on health spending, utilization, and health
Nishi et al. [45]	2012	Bulletin of the WHO	Health effects of patient cost-sharing
Pierce et al. [46]	2012	Pers Soc Psych Bulletin	Effect of income disparity in marriage
Sloan and Hanrahan [47]	2014	JAMA Ophthalmology	Effect of new therapies on vision loss among elderly patients
Smith et al. [48]	2014	Canadian Medical Association Journal	Effect of HPV vaccine on sexual behavior
Sood et al. [49]	2014	BMJ	Effect of health insurance on mortality
Weaver et al. [50]	2010	Journal of Traumatic Stress	Effect of cognitive-behavioral therapy on trauma symptoms
Yörük and Yörük [51]	2012	Social Science and Medicine	Effect of alcohol on psychological well-being

Regression Discontinuity Designs in Health

A Systematic Review

Michele Hilton Boon, Peter Craig, Hilary Thomson, Mhairi Campbell, and Laurence Moore

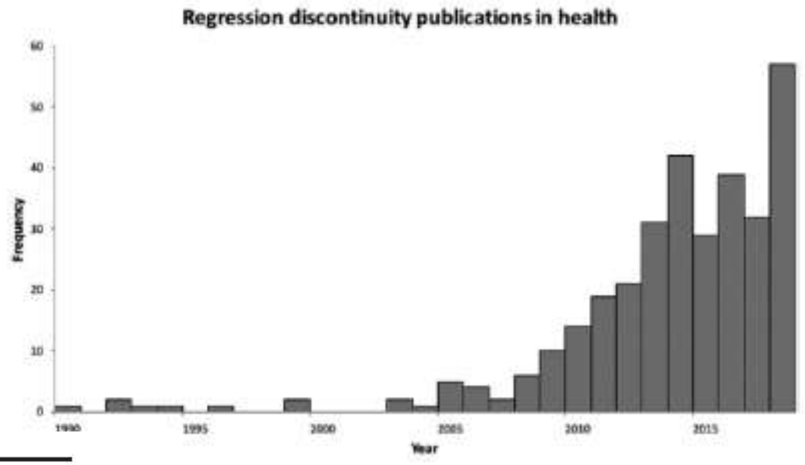


TABLE. Thematic Analysis of Forcing Variables and Threshold Rules Used in Regression Discontinuity Studies of Health Outcomes

Type of Forcing Variable	Number of Studies	Measurement Used	Threshold Rule
Age	110	Age in days, months, weeks, or years	Age threshold for: • Starting school • Leaving school • Legal drinking age • Treatment or benefit eligibility • Insurance eligibility • Retirement
Date/time	107	Calendar date, month, or year Time in minutes, hours, or days	Date or time of: • Implementation of policy/ legislation • Repeal of policy/legislation • Disaster or major incident • Change in situation or conditions
Socioeconomic measure	57	Company payroll total Dropout risk score Family income Household acreage Investment cost Poverty or literacy rate Poverty or welfare index Predicted probability of borrowing microcredit Program quality score Vote share or margin	Benefit or program eligibility Election outcome Legislated threshold

TABLE. Thematic Analysis of Forcing Variables and Threshold Rules Used in Regression Discontinuity Studies of Health Outcomes

Type of Forcing Variable	Number of Studies	Measurement Used	Threshold Rule
Clinical measure	31	Addiction severity measure	Risk threshold for intervention
		Birthweight	Guideline threshold for intervention
		Blood lead levels	Legislated threshold for intervention
		Body mass index	
		Cardiovascular risk	
		CD4 count	
		Down syndrome risk	
		Exeter Alcohol Scale	
		Hospital Safety Score	
		Parity	
		Positive Symptoms Scale	
		Posttraumatic Stress Disorder Reaction Index	
		Staffing numbers	
		Systolic blood pressure	
		Time of birth	
		Visual acuity	
		Weeks of gestation	
Environmental measure	6	Ozone forecasts	Policy threshold for action
		Air pollution levels	Legislated threshold for action
Geographical location	9	Political boundary	Program eligibility
		Distance from boundary	
		Latitude and longitude	
Other	6	Class size	Policy threshold for intervention/ exposure
		Number of schools	
		School test score	Program eligibility
		Village population	
		Draft lottery number	

A SIMPLE STEP-BY-STEP GUIDE TO RDD IMPLEMENTATION (MOSCOE ET AL. 2015)

- STEP 1: Determine feasibility of RD design
  - Continuous eligibility measure
  - Universal outcome assessment
  - Precise treatment assignment
- STEP 2: Consider covariate balance
  - To confirm that the groups on either side of the cutoff are comparable and that other observed factors are not discontinuous at the cutoff
- STEP 3: Visually check for a treatment effect
- STEP 4: Fit the regression models to estimate the treatment effect

FIVE ELEMENTS OF GOOD RD PRACTICE

1. A discussion of the **RD validity conditions** in the context of the particular study.
2. A clear presentation or discussion of the **assignment rule**.
3. **Covariate balance tests** for treated and non-treated groups showing that there is no discontinuity in pretreatment variables.
4. A histogram (frequency) of the assignment variable that **shows no bunching of the data around the cutoff** to demonstrate no manipulation of treatment status.
5. **Multiple RD estimation specifications** to check for robustness, including alternative functional forms, nonparametric regression, etc …

INCLUSION OF KEY ELEMENTS IN RDD STUDIES

Author	Year	Discussion of RD validity conditions	Discussion of assignment rule	Covariate balance tests	Histogram of assignment variable	Robustness checks	Total (0-5)
Albouy and Lequien [23]	2009	✓	✓	✗	✗	✓	3
Almond et al. [15]	2010	✓	✓	✓	✓	✓	5
Andalón [24]	2011	✓	✓	✓	✓	✓	5
Anderson et al. [25]	2011	✓	✓	✗	✗	✗	2
Arcand and Wouabe [26]	2010	✓	✓	✓	✗	✓	4
Banks and Mazzonna [27]	2012	✓	✓	✓	✗	✓	4
Behrman [28]	2014	✓	✓	✗	✗	✓	3
Bor et al. [1]	2014	✓	✓	✓	✓	✓	5
Callaghan et al. [29]	2014	✓	✓	✗	✗	✓	3
Callaghan et al. [30]	2013	✓	✓	✓	✗	✓	4
Callaghan et al. [31]	2013	✗	✓	✗	✗	✓	2
Carpenter and Dobkin [16]	2009	✓	✓	✓	✗	✓	4
Carpenter and Dobkin [32]	2011	✓	✓	✗	✗	✗	2
Chen et al. [33]	2013	✓	✓	✓	✓	✓	5
Conover and Scrimgeour [34]	2013	✗	✓	✗	✗	✓	2
De La Mata [35]	2012	✓	✓	✓	✗	✓	4
Deza [36]	2014	✓	✓	✓	✗	✓	4
Flam-Zalcman et al. [37]	2012	✓	✓	✗	✗	✗	2
Fletcher [38]	2014	✓	✓	✓	✓	✓	5
Glance et al. [39]	2014	✗	✓	✓	✗	✓	3
Gormley et al. [40]	2005	✓	✓	✓	✗	✓	4
Huang and Zhou [41]	2013	✓	✓	✗	✗	✓	3
Jensen and Wust [42]	2014	✓	✓	✓	✗	✓	4
McFarlane et al. [43]	2014	✓	✓	✓	✗	✓	3
Miller et al. [44]	2013	✓	✓	✓	✓	✓	5
Nishi et al. [45]	2012	✗	✓	✗	✗	✗	2
Pierce et al. [46]	2012	✓	✓	✓	✗	✓	5
Sloan and Hanrahan [47]	2014	✗	✓	✓	✗	✗	2
Smith et al. [48]	2014	✓	✓	✓	✗	✓	4
Sood et al. [49]	2014	✓	✓	✓	✗	✓	4
Weaver et al. [50]	2010	✗	✓	✓	✗	✗	3
Yörük and Yörük [51]	2012	✓	✓	✓	✗	✓	4

FURTHER CONSIDERATIONS IN THE REGRESSION DISCONTINUITY DESIGN

- The estimand of interest is the Local Average Treatment Effect (LATE)
  - AKA complier average causal effect (CACE)
  - Defined in parallel in econometrics and public health in the 80-90s
    - Baker, Stuart G., and Karen S. Lindeman. "Multiple Discoveries in Causal Inference: LATE for the Party." CHANCE 37.2 (2024): 21-25.
  - External validity challenges
- When the probability of receiving the treatment is non-deterministic, fuzzy regression discontinuity designs can be employed
  - Example: the Toronto Air Quality Program
- In summary, an ideal RDD setting, like for RCTs, can rely on very simple techniques, but in practice, it is a different story, but various solutions exist

Effect of air quality alerts on health outcomes in Toronto, Canada: a regression discontinuity analysis

Hong Chen, Qiongsi Li, Jay S Kaufman, Jun Wang, Ray Copes, Yushan Su, Tarik Benmarhnia

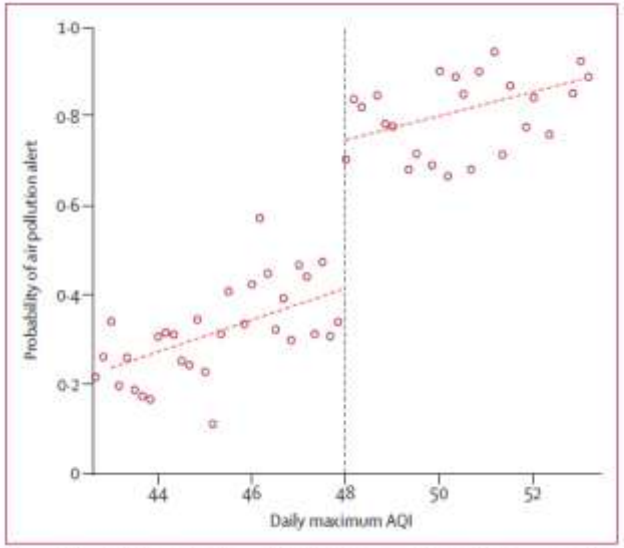


Figure 1: Probability of an air quality alert in Toronto, Canada, from 2003 to 2012, by daily maximum AQI
Plot is centred at an AQI of 48 (blue vertical dotted line). Eligible days were days with daily maximum AQI ≥48 (dots to the right) and non-eligible days were days with daily maximum AQI <48 (dots to the left). Red dashed lines depict linear regression lines. AQI=air quality index.

REGRESSION DISCONTINUITY WITH TIME AS THE FORCING VARIABLE

- Several studies used discontinuity in time as their identification strategy
- Example: Protective Effects of Smoke-free Legislation on Birth Outcomes in England

Protective Effects of Smoke-free Legislation on Birth Outcomes in England

A Regression Discontinuity Design

Ioannis Bakolis,[a,b,c] Ruth Kelly,[d,e] Daniela Fecht,[a] Nicky Best,[d] Christopher Millett,[f] Kevin Garwood,[a] Paul Elliott,[a,g] Anna L. Hansell,[a,g] and Susan Hodgson[h]

- In this setting, a regression discontinuity design is essentially the same as interrupted time series (ITS) analyses
  - Which is like an uncontrolled difference-in-differences

Annual Review of Resource Economics

Regression Discontinuity in Time: Considerations for Empirical Applications

Catherine Hausman[1] and David S. Rapson[2]

- In their review, Hausman and Rapson, discuss several features of the regression discontinuity in time framework
- They highlight some potential sources of biases including:
  - Time-Varying Treatment Effects
  - Autoregression and seasonal patterns
  - Anticipation, lagged effects and harvesting effects
- These issues are well documented in the ITS literature
- The modelling approach in a typical RD analysis and a 1-stage ITS are the same
  - Modelling time trends (long term, seasonal..) and consideration of time-varying covariates
  - Modelling changes in the intercept and the slope

USING COVARIATES IN THE RDD

- Some RDD studies also rely on covariates
  - Mostly for falsification tests: placebo outcomes
  - Covariate adjustment can be used to improve precision in estimation and inference
- Yet, it is also possible to consider placebo outcomes as control units
- For example, consider an example focusing on minimum drinking age and age as the forcing variable.
- It is possible to use (for example) multiple fruits and vegetables consumption as control variables and possibly build a synthetic control group
  - Sofer, Tamar, et al. "On negative outcome control of unobserved confounding as a generalization of difference-in-differences." Statistical science: a review journal of the Institute of Mathematical Statistics 31.3 (2016): 348.

Annual Review of Economics

Regression Discontinuity Designs

Matias D. Cattaneo[1] and Rocío Titiunik[2]

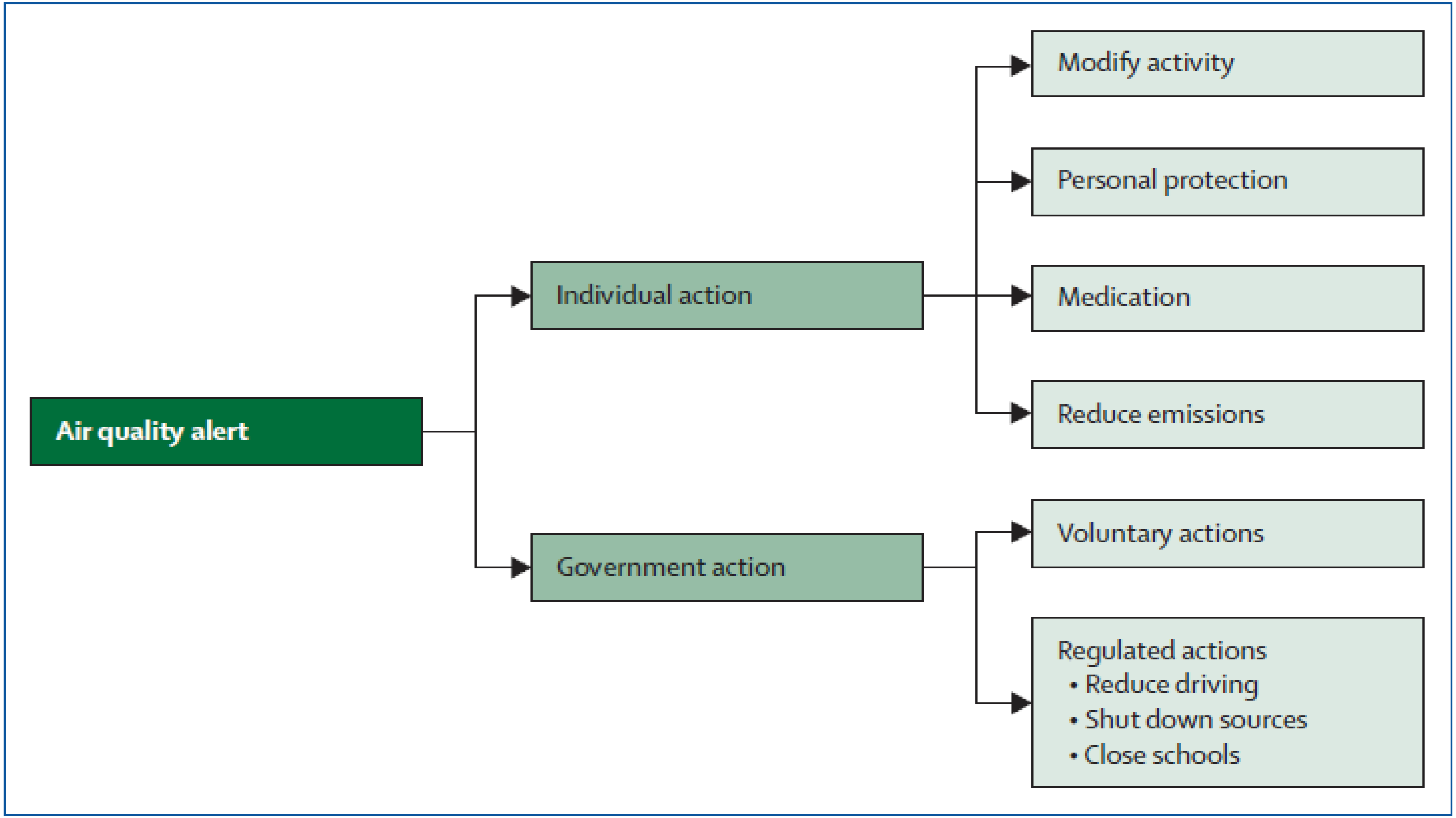
<u>Some useful resources</u>

- Calonico S, Cattaneo MD, Farrell MH, Titiunik R. 2019. Regression discontinuity designs using covariates. Rev. Econ. Stat. 101(3):442–51
- Cattaneo MD, Keele L, Titiunik R. 2022. Covariate adjustment in regression discontinuity designs. In Handbook of Matching and Weighting in Causal Inference, ed. JR Zubizarreta, EA Stuart, DS Small, PR Rosenbaum. London: Chapman & Hall

AIR QUALITY ALERTS

- Many cities worldwide have implemented Air Quality Alerts
  - aka Smog Warning systems
- Activated during Smog Days
  - days with elevated levels of air pollution

HOW DO AIR QUALITY ALERTS WORK?



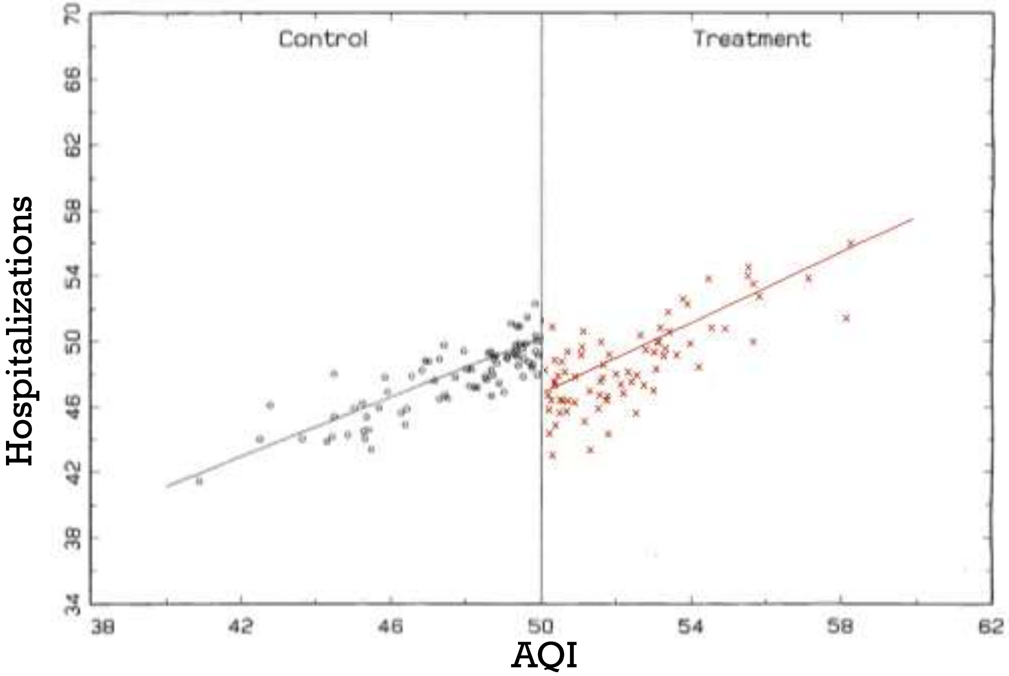
THE CONTEXT IN TORONTO

- Toronto is among the major North American cities that most frequently have air quality alerts
- The air quality alert in Toronto **started in 2000**
  - Reduce outdoor activities
  - Reduce emissions (speed limitations, shut down industries…)
  - Close schools
  - …
- Air quality alert are issued based on a threshold of the air quality index (AQI):
  - AQI includes PM2.5, NO2 and O3
- **Air Quality Alert triggered when AQI ≥50**

Effect of air quality alerts on health outcomes in Toronto, Canada: a regression discontinuity analysis

Hong Chen, Qiongsi Li, Jay S Kaufman, Jun Wang, Ray Copes, Yushan Su, Tarik Benmarhnia

- The relationship between air pollution and health is essentially linear or log-linear
- The threshold used for triggering air quality alerts (i.e. AQI ≥50) <u>does not correspond</u> to any known biological mechanism
- Days <u>**slightly above and below**</u> this AQI threshold are expected to be **very similar** (i.e. exchangeable) <u>without the policy</u>



MAIN RESULTS

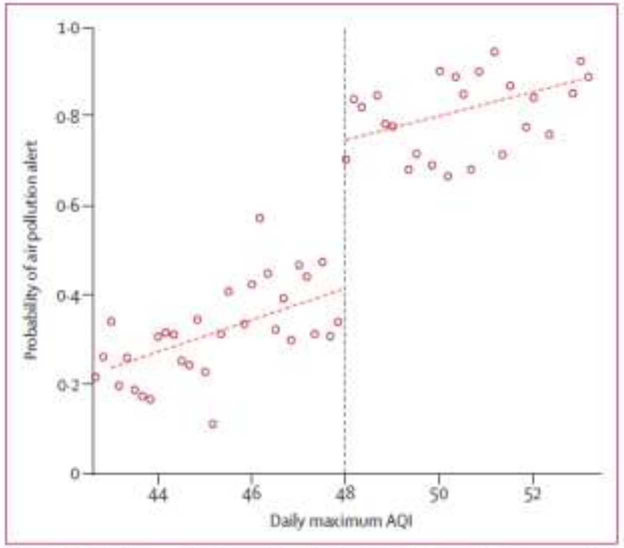


Figure 1: **Probability of an air quality alert in Toronto, Canada, from 2003 to 2012, by daily maximum AQI**
Plot is centred at an AQI of 48 (blue vertical dotted line). Eligible days were days with daily maximum AQI ≥48 (dots to the right) and non-eligible days were days with daily maximum AQI <48 (dots to the left). Red dashed lines depict linear regression lines. AQI=air quality index.

Panel: Regression models

Equation 1

$$g\{E(Y_i)\}=\beta_0+\beta_1\times(Z_i-c)+\beta_2\times P_i+\beta_3\times(Z_i-c)\times P_i+\beta_k\times X_{ik}$$

Equation 2

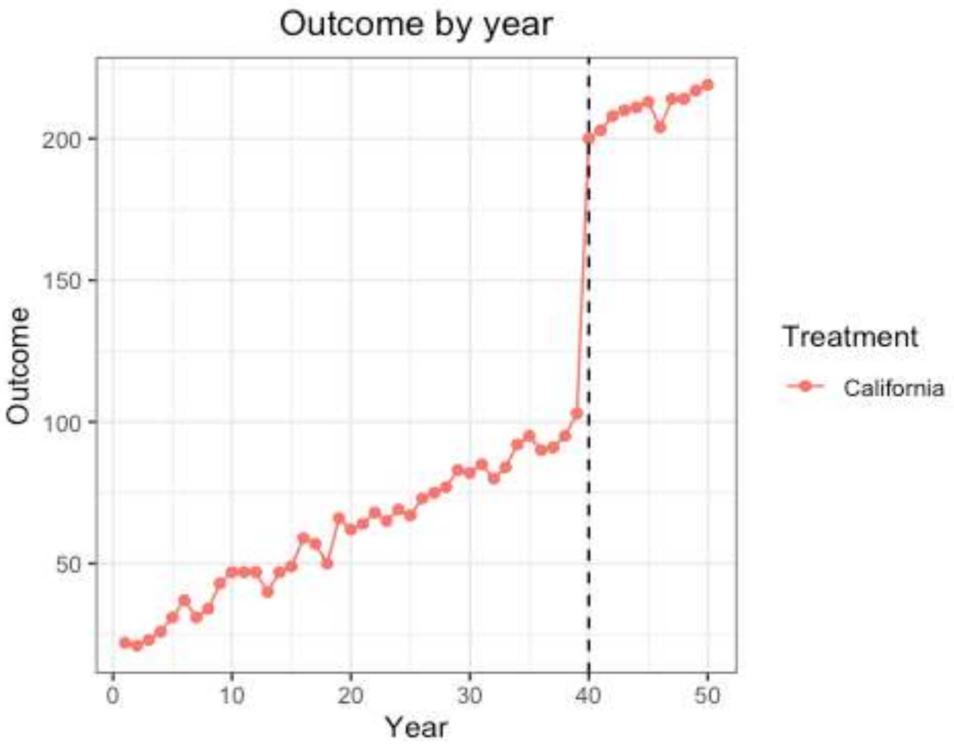
$$g\{E(T_i)\}=\delta_0+\delta_1\times(Z_i-c)+\delta_2\times P_i+\delta_3\times(Z_i-c)\times P_i+\delta_k\times X_{ik}$$

$g\{.\}$ is a generic link function. i denotes day i between 2003 and 2012. Y_i is the daily count of selected outcomes. Z_i is daily air quality index with a threshold at c (ie, 48). P_i is an indicator variable that reflects eligibility status (1 for day i with $Z_i \geq c$, otherwise 0). T_i is an indicator variable equal to 1 for day i with an alert. X_{ik} are a set of k covariates including temperature (using a natural spline with three degrees of freedom), relative humidity, calendar year, season, weekend, and holidays.

BREAK

NATURAL EXPERIMENTS BASED ON TIMING, ADVANCED TOPICS

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

$$y = \alpha + \beta_1(\text{year}) + \beta_2(\text{post}) + \beta_3(\text{xi}) + \beta_4(\text{xt}) + \beta_5(\text{xit}) + \beta_6(\text{year} \times \text{post}) + \epsilon$$

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)

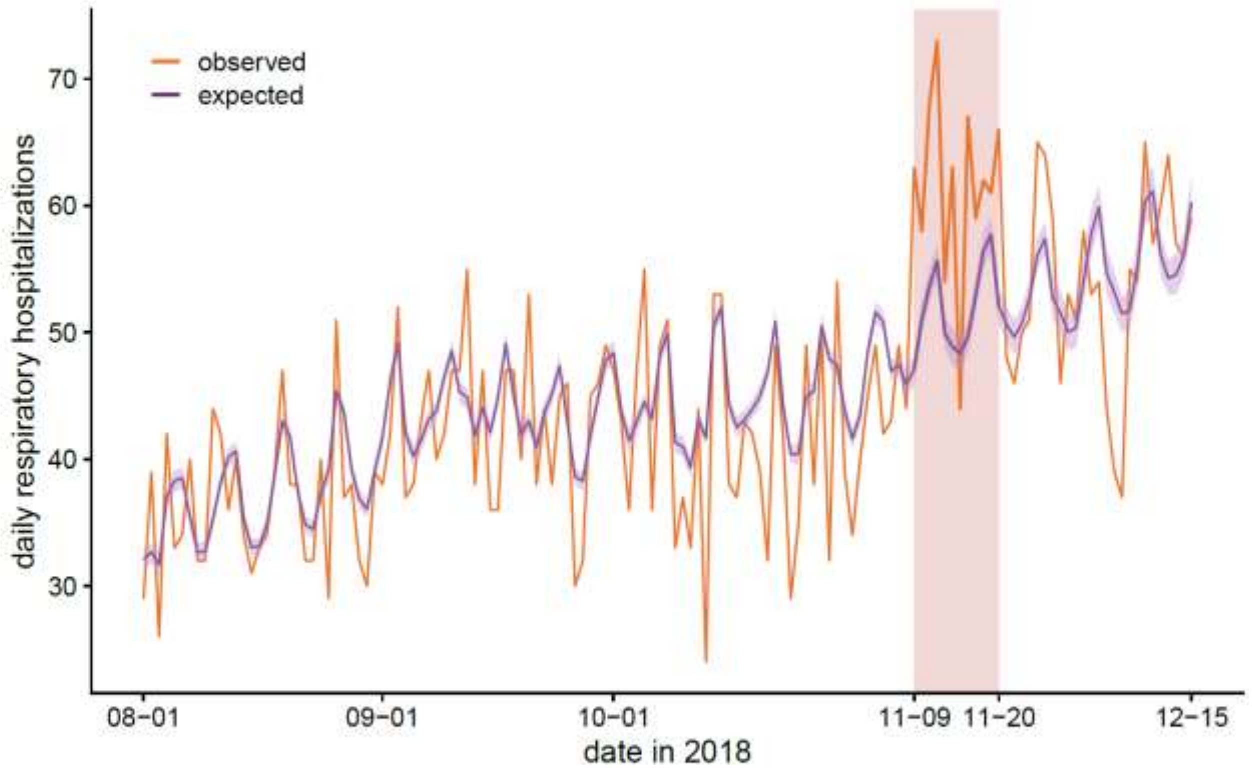
TWO-STAGE TIME SERIES ANALYSIS COUPLED WITH MACHINE LEARNING: EVALUATING THE HEALTH EFFECTS OF THE 2018 WILDFIRE SMOKE EVENT IN SAN FRANCISCO COUNTY AS A CASE STUDY

- 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

R^2: coefficient of determination
MAE: mean absolute error
RMSE: root mean square error
MAPE: mean absolute percentage error
SMAPE: symmetric mean absolute percentage error



OTHER ITS TOPICS NOT DISCUSSED TODAY

- Using Bayesian ITS
- Staggered Interventions

Florida's Opioid Crackdown and Mortality From Drug Overdose, Motor Vehicle Crashes, and Suicide: A Bayesian Interrupted Time-Series Analysis

Kenneth A. Feder, Ramin Mojtabai, Elizabeth A. Stuart*, Rashelle Musci, and Elizabeth J. Letourneau

Education Corner

Staggered interventions with no control groups

Brice Batomen[1,*] **and Tarik Benmarhnia**[2,3]

[1]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, [2]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA and [3]Irset Institut de Recherche en Santé, Environnement et Travail, Inserm, University of Rennes, EHESP, Rennes, France

*Corresponding author. Department of Epidemiology, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Room 688, Toronto, ON M5T 3M7, Canada. E-mail: brice.kuimi@utoronto.ca

Key Messages

- In staggered intervention scenarios without control groups, common model specification for impact evaluations may yield biased estimates due to ill-defined post-intervention periods.
- Alternative model specifications that are drawn from the difference-in-differences literature for staggered interventions can be easily adopted when no control group is available.
- The adoption of these alternative models improves the validity of impact evaluations, especially if heterogeneity is expected across treated groups and across post-intervention time periods.

MULTIPLE TREATED GROUPS

- When dealing with multiple treated units, there are two distinct settings:
  - All treated units received the intervention of interest at the same time
  - Treated units received the intervention at different times
- When multiple units receive the intervention at the same time:
  - A simple approach is to estimate a separate ATT for each treated unit and then conduct a meta-analysis to get a pooled estimate (and also information about heterogeneity across units)
  - Or aggregate the treated units and form a synthetic control for the aggregated treated unit (Acemoglu et al. 2013; Dube and Zipperer, 2015)

MULTIPLE TREATED UNITS AT DIFFERENT TIMES

- A very active area of research
  - Goodman-Bacon (2018) proposed a solution based on a weighted average of all possible DID estimators (using some groups multiple times) in the sample of interest. This approach requires an additional identifying assumption of time-invariant treatment effects.
  - Callaway and Sant'Anna (2020) proposed an analytical solution for such case where there are more than two time periods and units that can become treated at different points in time while relaxing the time-invariant treatment effects assumption

WHAT DO WE MEAN BY STAGGERED INTERVENTIONS?

- When a given policy/treatment is affecting multiple units <u>but at different times points</u>

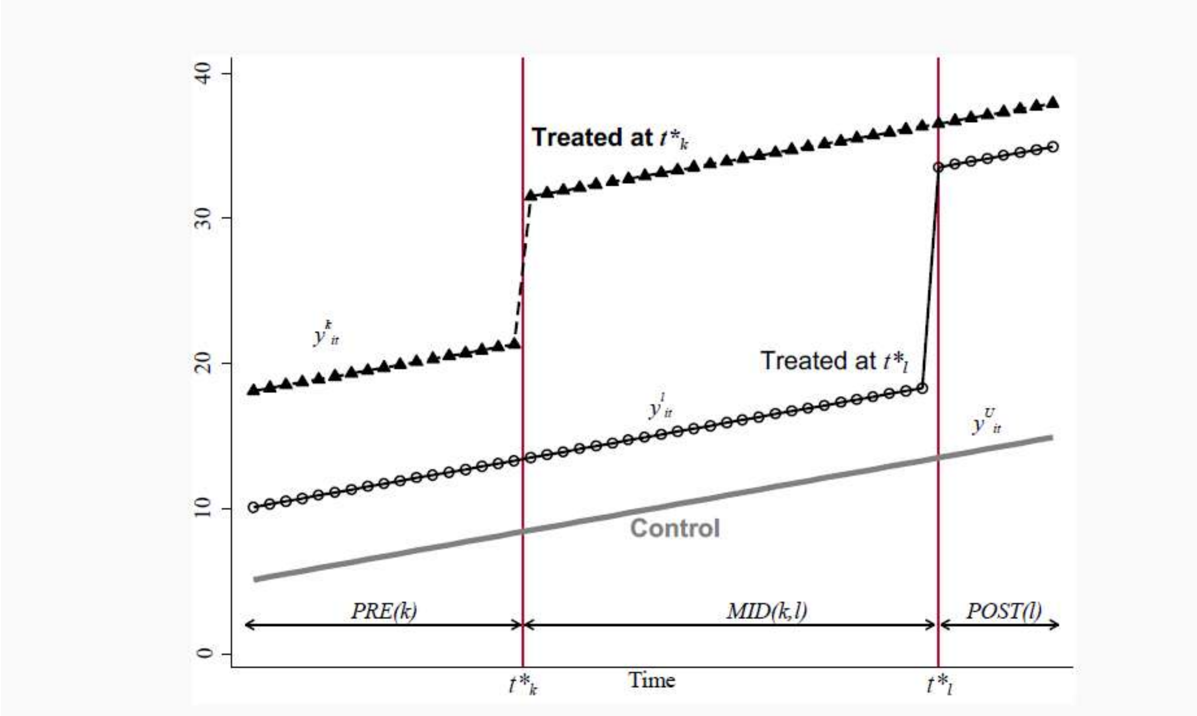


Figure from Goodman-Bacon, 2021

WHY DO TRADITIONAL METHODS FAIL IN SUCH SETTINGS?

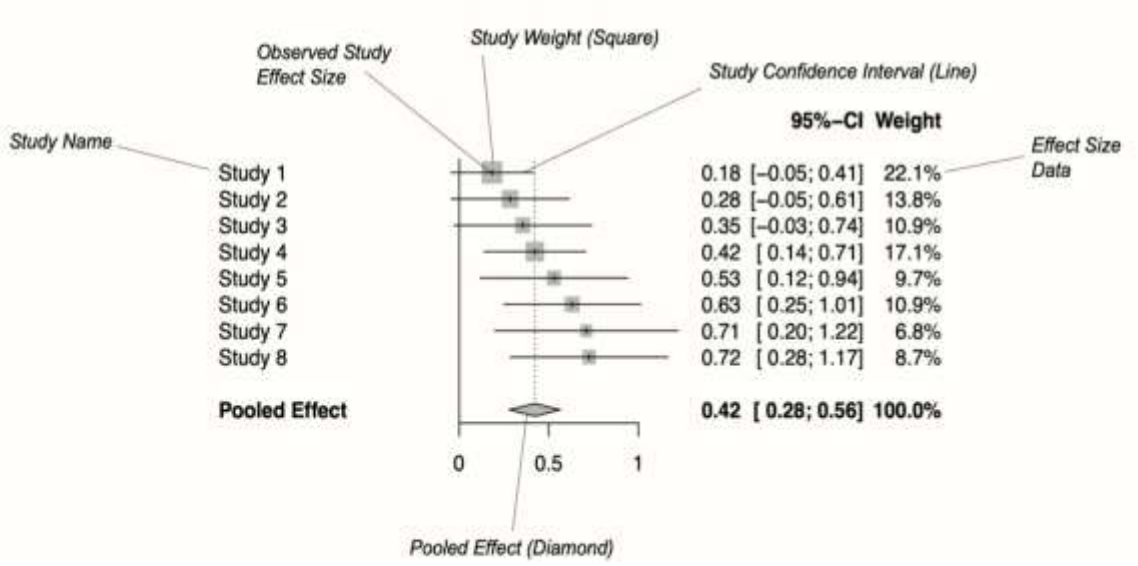
- Simple settings for DID and SCM do not apply here for a few reasons:
  - Each treated unit requires a suitable control group specific the timing of the policy/treatment
  - Units that receive the policy/treatment later may be used as control unit earlier in the study period
    - Some potential statistical dependence issues
  - **We do not want to assume that a given policy/treatment has the same effect across units and over time**

ANALYTICAL APPROACHES TO DEAL WITH STAGGERED INTERVENTIONS

- Multiple approaches have been proposed in the past few years
  - A very active area: De Chaisemartin, C., & d'Haultfoeuille, X. (2022)
- We follow the 2-stage approach proposed by Callaway and Sant'Anna (2020)
- This approach requires:
  - To have –enough- never treated units
  - To assume independence regarding the timing of the implementation over treated units
  - Similar assumptions as traditional settings

A QUICK NOTE ON META-ANALYSES AND META-REGRESSIONS

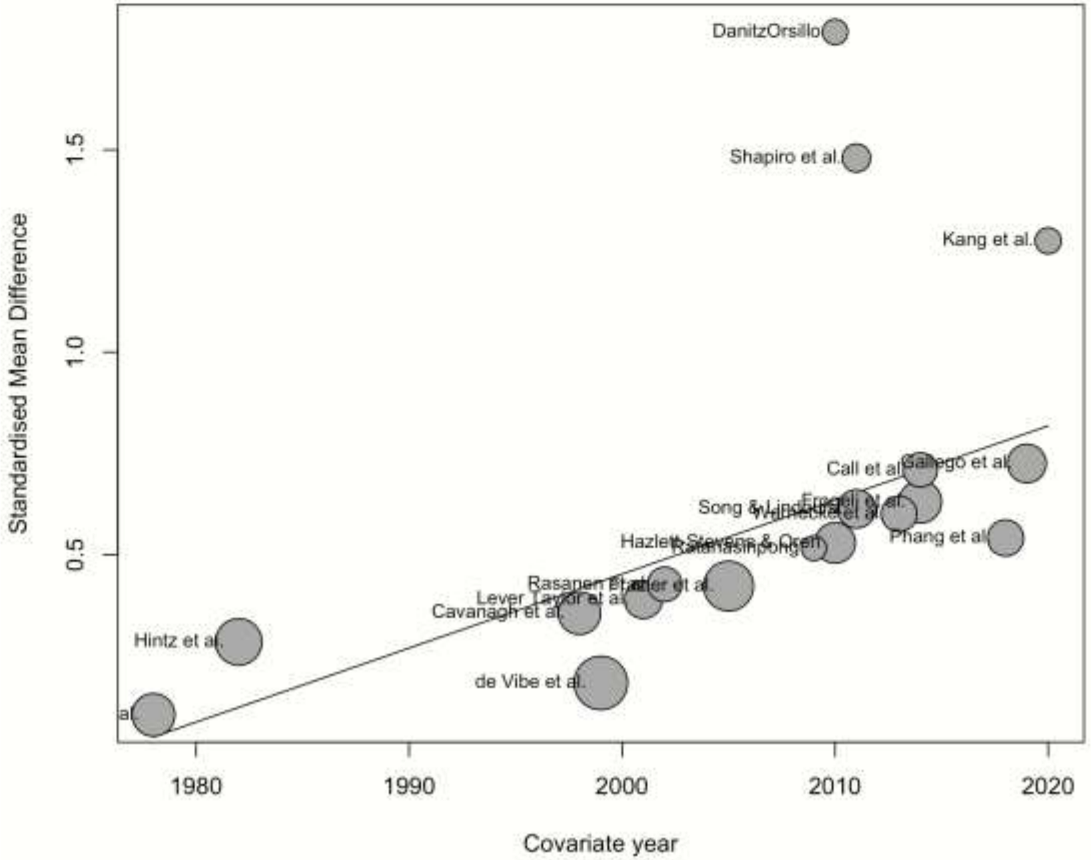
- Meta-analyses are typically used to pool effect estimates from multiple studies
- This is a form of multilevel model, in which participants are nested within studies
- We obtain a pooled effect estimate by applying a weighting procedure (usually based on the inverse of the standard error)
- We can also quantify the level of heterogeneity across studies
- <u>If there is some heterogeneity</u>, we then can conduct a meta-regression in which the dependent variable is the study-specific effect estimate and independent variables study-specific characteristics (e.g. study period, population composition etc…)



https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/forest.html

HOW DO META-REGRESSION WORK?

- An intuitive way to describe meta-regression is a weighted regression based on the study-specific variance
  - If all studies have the same variance: meta-regression = simple regression
- We then obtain a slope for each independent variable, and we can understand the drivers of effect estimates heterogeneity

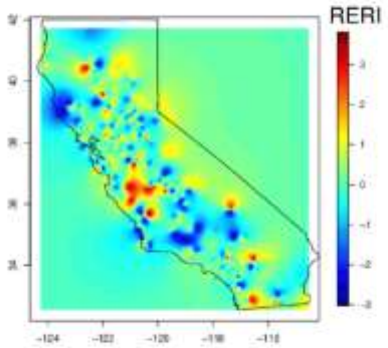


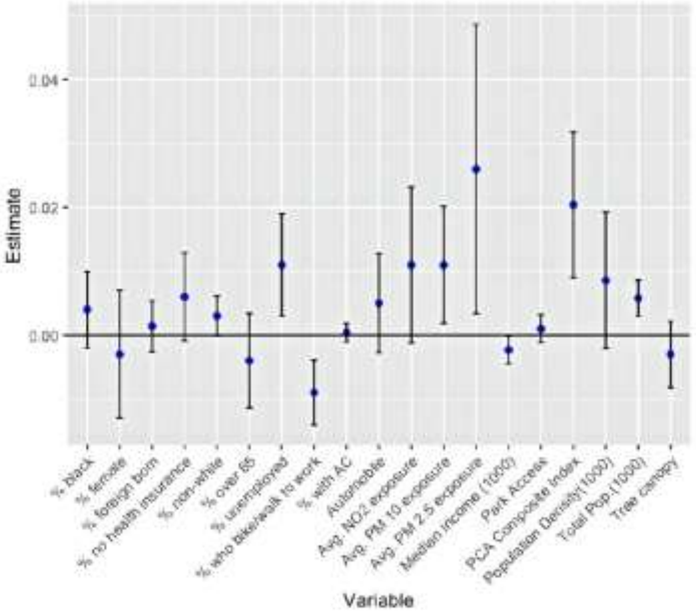
https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/metareg.html

SOME EXAMPLES

Spatial variation in the joint effect of extreme heat events and ozone on respiratory hospitalizations in California

Lara Schwarz[a,b,1,2], Kristen Hansen[b,1,2], Anna Alari[c], Sindana D. Ilango[d], Nelson Bernal[e], Rupa Basu[f], Alexander Gershunov[g], and Tarik Benmarhnia[b,g]





Epidemiology of *Chlamydia trachomatis* in the Middle East and north Africa: a systematic review, meta-analysis, and meta-regression

Alex Smolak*, Hiam Chemaitelly*, Joumana G Hermez, Nicola Low, Laith J Abu-Raddad

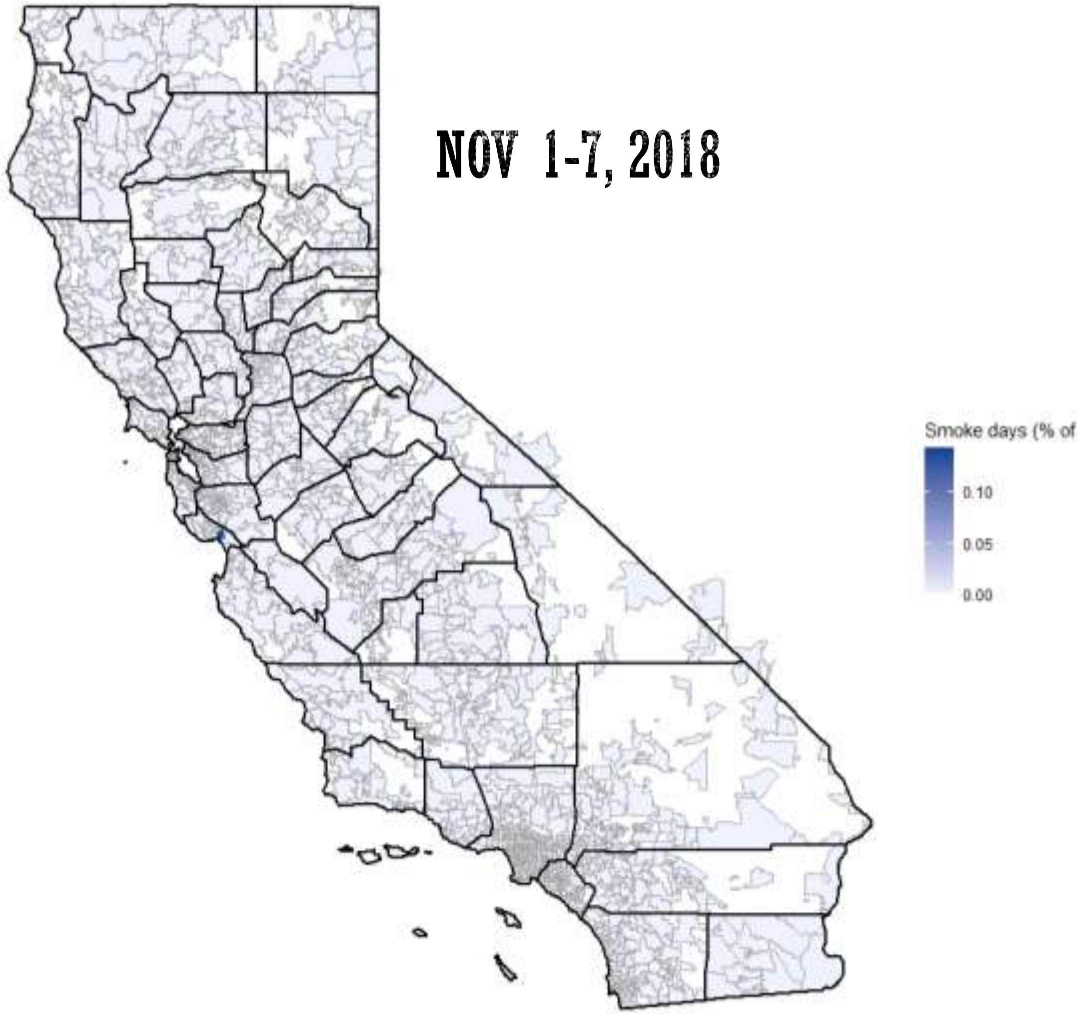
Sex differences in injury rates in team-sport athletes: A systematic review and meta-regression analysis

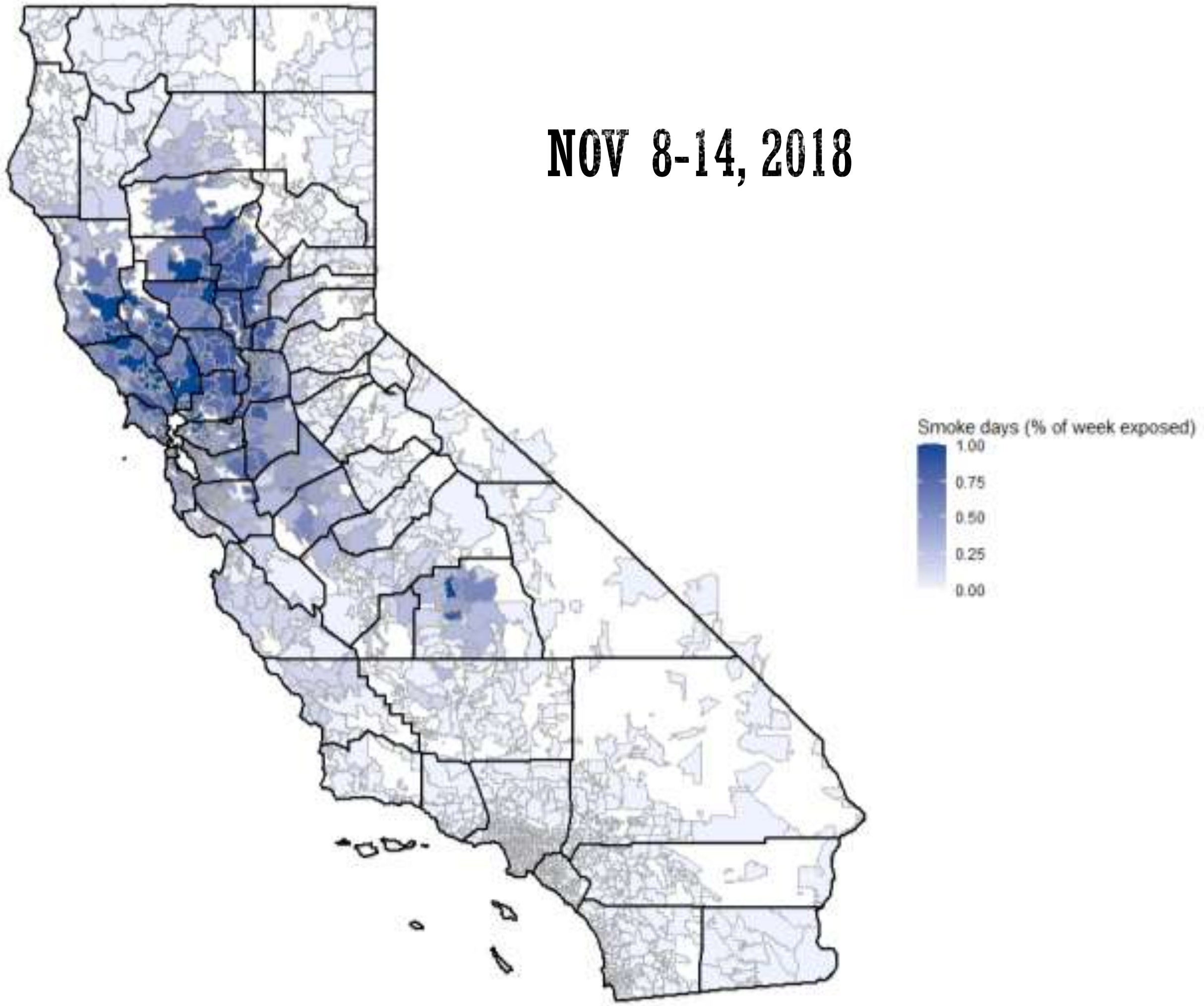
Astrid Zech[a,*], Karsten Hollander[b], Astrid Junge[b,c], Simon Steib[d], Andreas Groll[e], Jonas Heiner[e], Florian Nowak[a], Daniel Pfeiffer[a], Anna Lina Rahlf[f]

A CASE STUDY:
WILDFIRE SMOKE AND COMPOUNDED RISKS FROM RESPIRATORY INFECTIOUS DISEASES

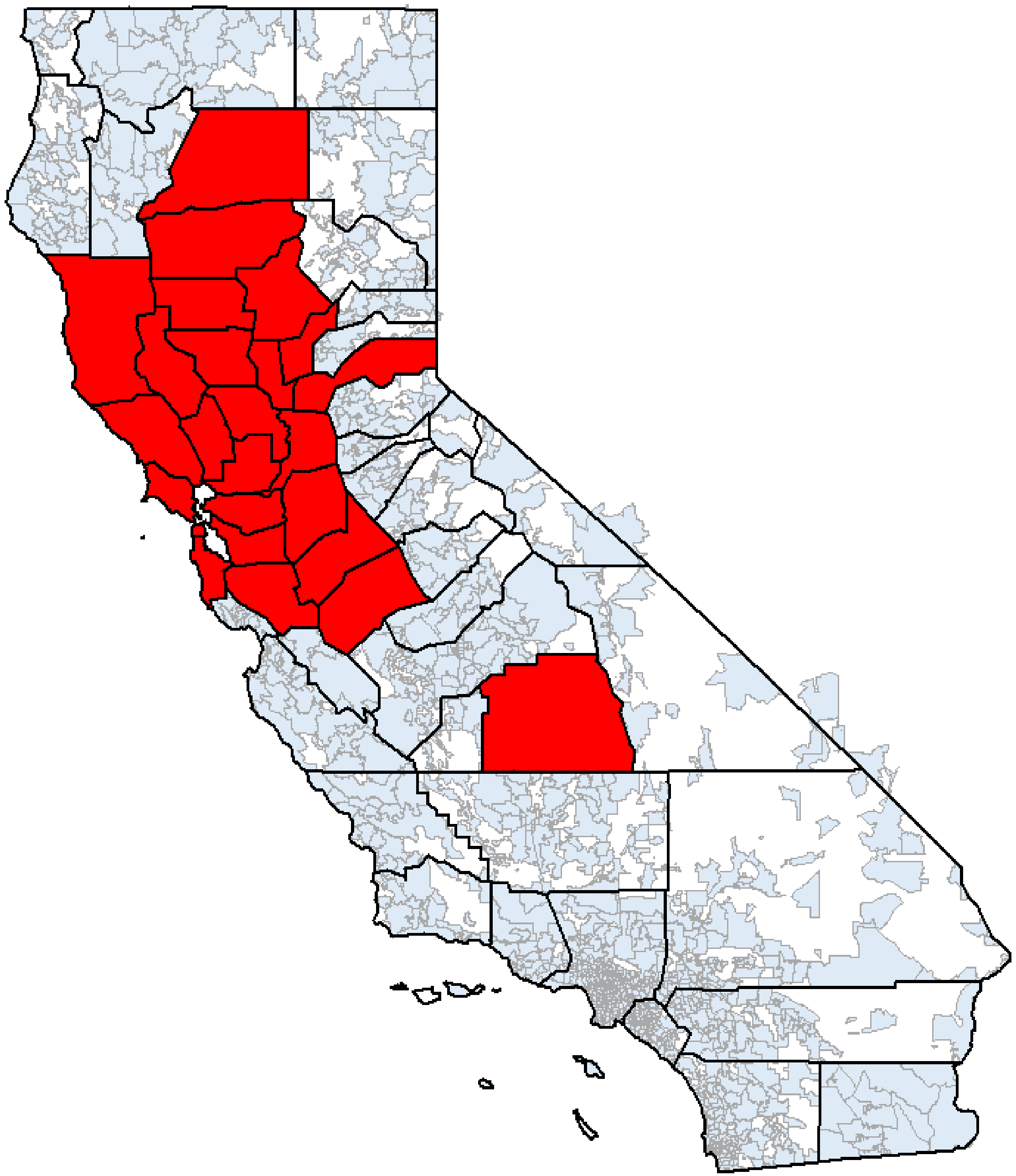
- Objectives:
  - Assess the county-specific effects of wildfire smoke on all respiratory hospitalizations using a large wildfire event in November 2018 as a case study
    - Using generalized synthetic control methods
  - Evaluate how population-level influenza frailty modifies such effects
    - Using a random effect meta-regression

TREATED ZIP CODES WITH SMOKE PM2.5>0





25 COUNTIES EXPOSED DURING NOV 8-14 WEEK



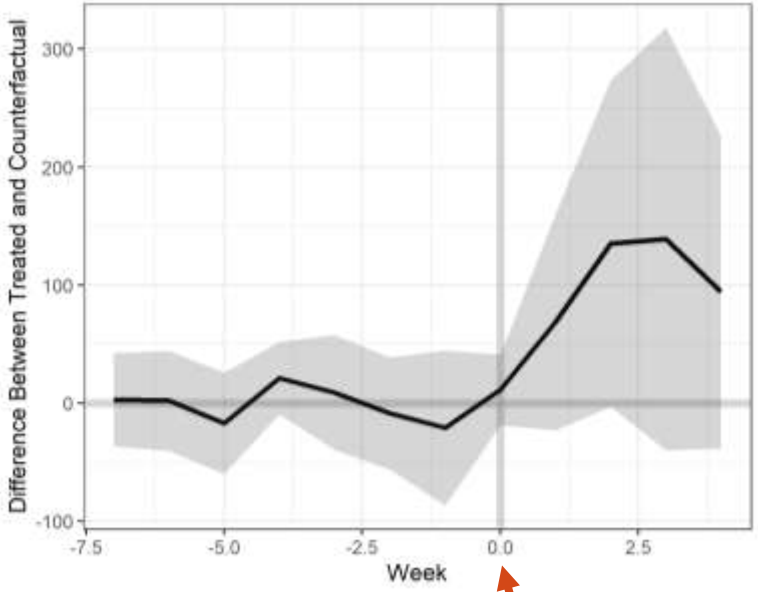
ESTIMATING INFLUENZA BURDEN Z-SCORE

- OSHPD flu data was processed at the daily zip code level
  - ICD-9 codes: 487, 488 ICD-10 codes: J09, J10, J11
- Weekly counts were summarized at the County level
- Weekly z-score computed for every County
  - Restricted to 2010-2019, flu season based on CDC (October to May)
- Average flu z-score for weeks 42-44 (**<u>3 weeks prior</u>** to wildfire smoke start) for each County

*OSHPD: Office of Statewide Health Planning and Development Patient Discharge Data

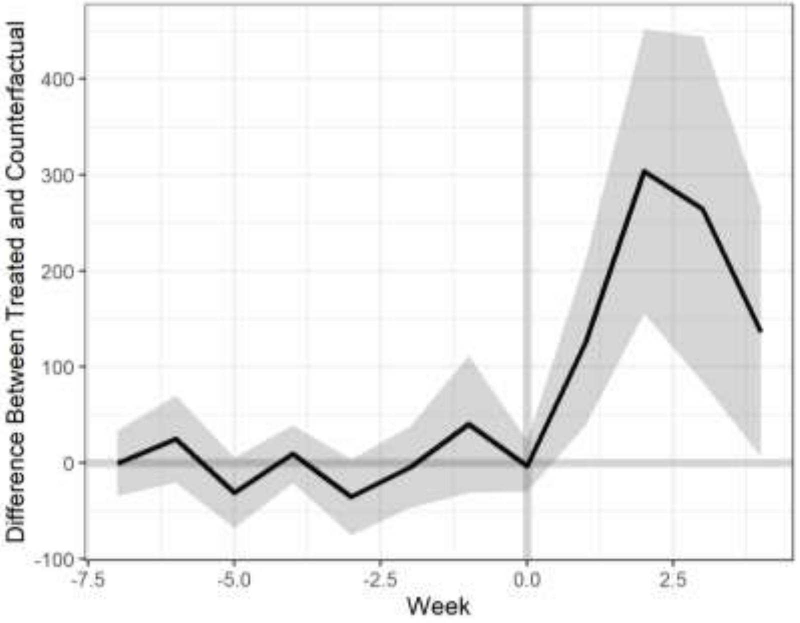
RESULTS (1/2)

SAN JOAQUIN COUNTY

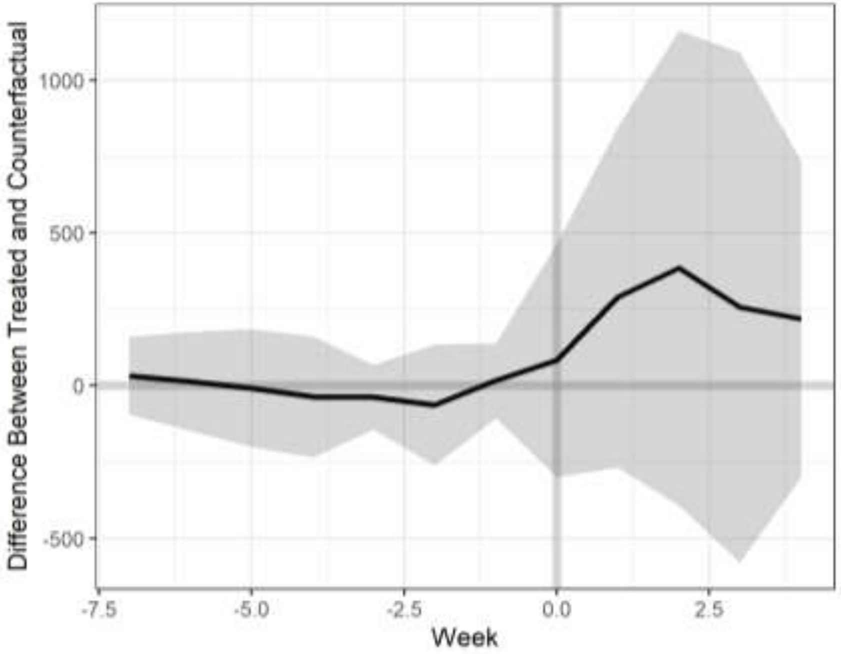


WILDFIRE START

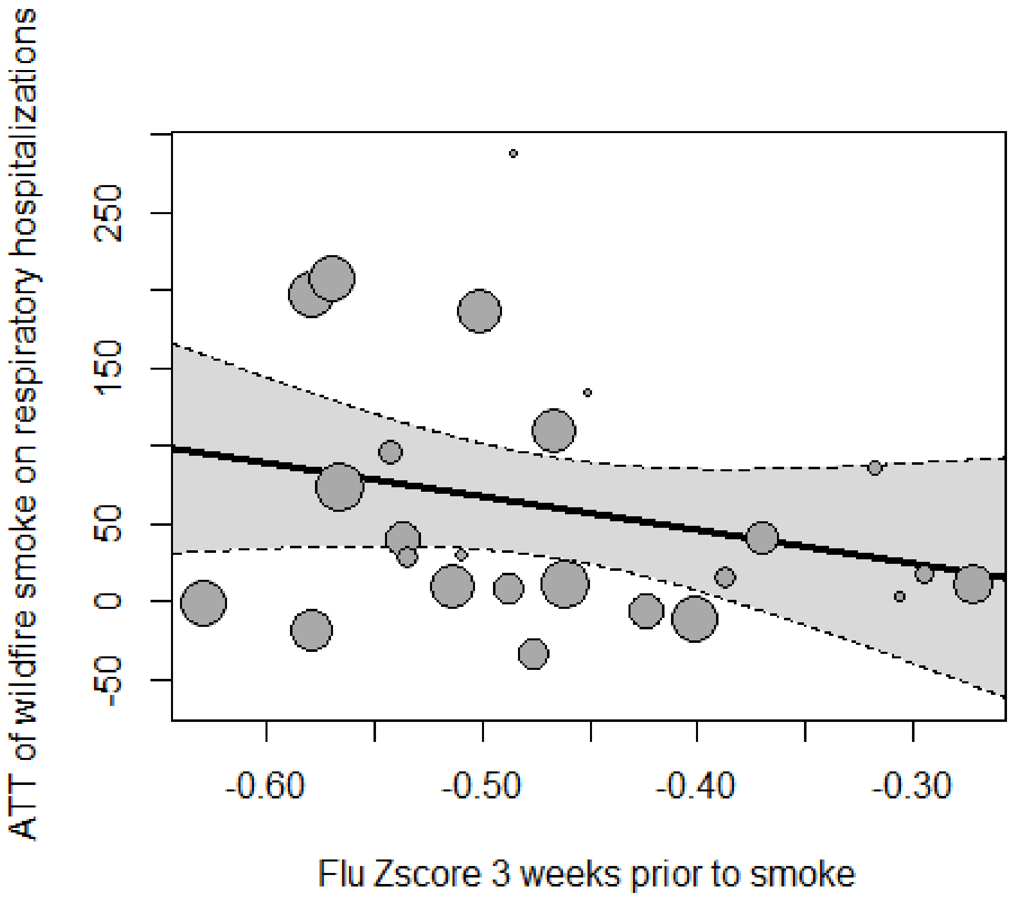
SANTA CLARA COUNTY



ALAMEDA COUNTY



RESULTS (2/2)



RESULTS FROM THE RANDOM-EFFECT META-REGRESSION MODEL

β = -215.42 (95% CI:, -550.72, -119.88)

THANK YOU

YIM022 @ UCSD.EDU
TBENMARHNIA @ UCSD.EDU