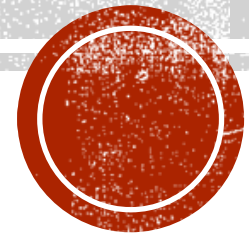# An introduction to Difference-in-Differences and Synthetic Control Methods for Epidemiologists

09 July 2021

SER 54th Annual Meeting Workshop

Roch Nianogo and Tarik Benmarhnia

AGENDA FOR TODAY

- Introductions

- Part I ~ 80 min

- Q&A

- Break

- Part II ~ 90-120 min

- Q&A

- End

OUTLINE – PART 1

- Causal inference, natural experiments and quasi-experimental methods

- Difference-in-differences methods

- Interrupted Time Series

- Synthetic Control Methods

- More advanced topics

WORKSHOP OBJECTIVES

1. Understand the conceptual foundations of quasi-experimental methods based on the timing of a natural experiment

2. Understanding how Difference-in-Differences (DID), Interrupted Time Series (ITS) and Synthetic Control Methods (SCM) work and their specific identification assumptions.

3. Being able to implement (in R) DID, ITS and SCM in different settings, interpret estimands of interest and check/visualize possible assumptions' violations

RANDOMIZATION AS A SOLUTION TO DEAL WITH CONFOUNDING

- Randomization has been proposed as a solution to deal with confounding
  - Complying with exchangeability and substitution between exposed and non-exposed to the policy/treatment
  - For both **measured & unmeasured** confounding

IDEAL RANDOMIZED EXPERIMENTS

- What makes an ideal randomized experiment:
  - No loss to follow-up,
  - Full adherence to the assigned treatment over the duration of the study,
  - A single version of treatment, and double-blind assignment
- **Ideal randomized experiments are unrealistic but useful to introduce some key concepts for causal inference**

THE IDEA BEHIND RANDOMIZATION

- How do RCT work …

- Randomization can be analyzed as an "**Instrumental Variable**"

- The overall aim is to deal with measured and unmeasured confounding in observational studies by using an instrumental variable
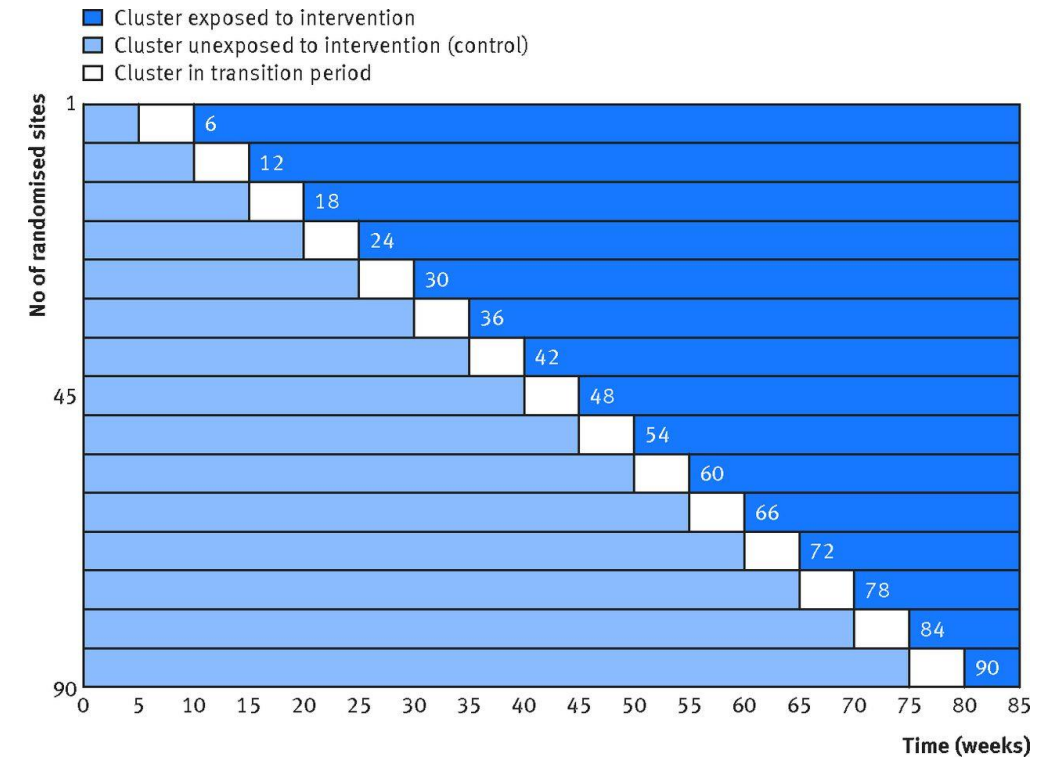

- **The problem:** in many cases RCT are not feasible due costs, scale or ethical objections

- 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**
    - **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.

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 deal with both measured and unmeasured confounding

EXAMPLE OF NATURAL EXPERIMENTS
NOT ONLY FOR POLICY EVALUATION

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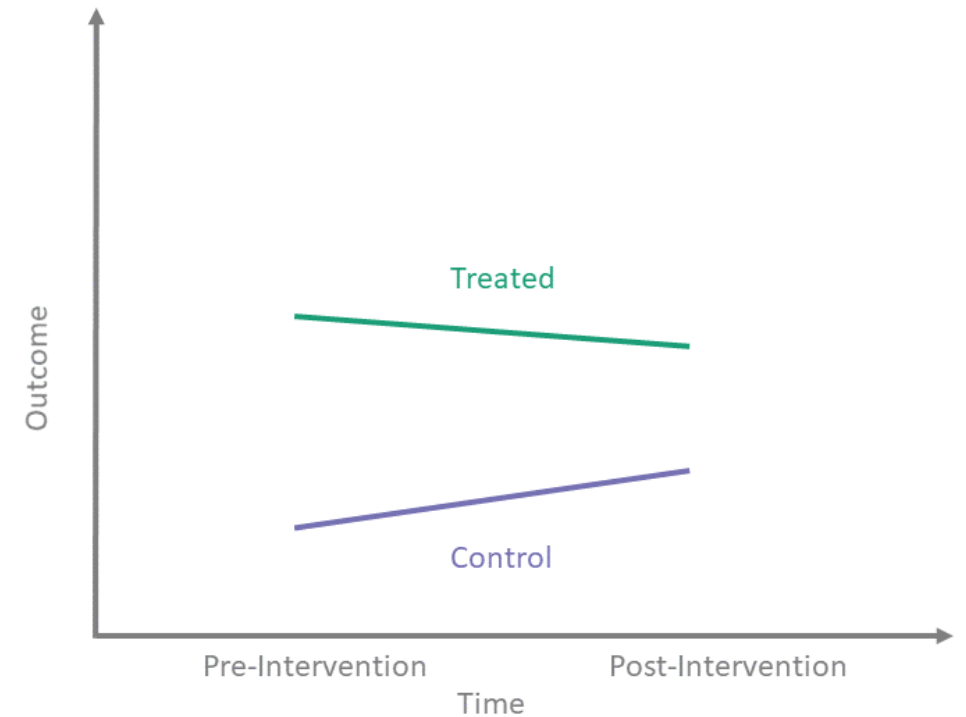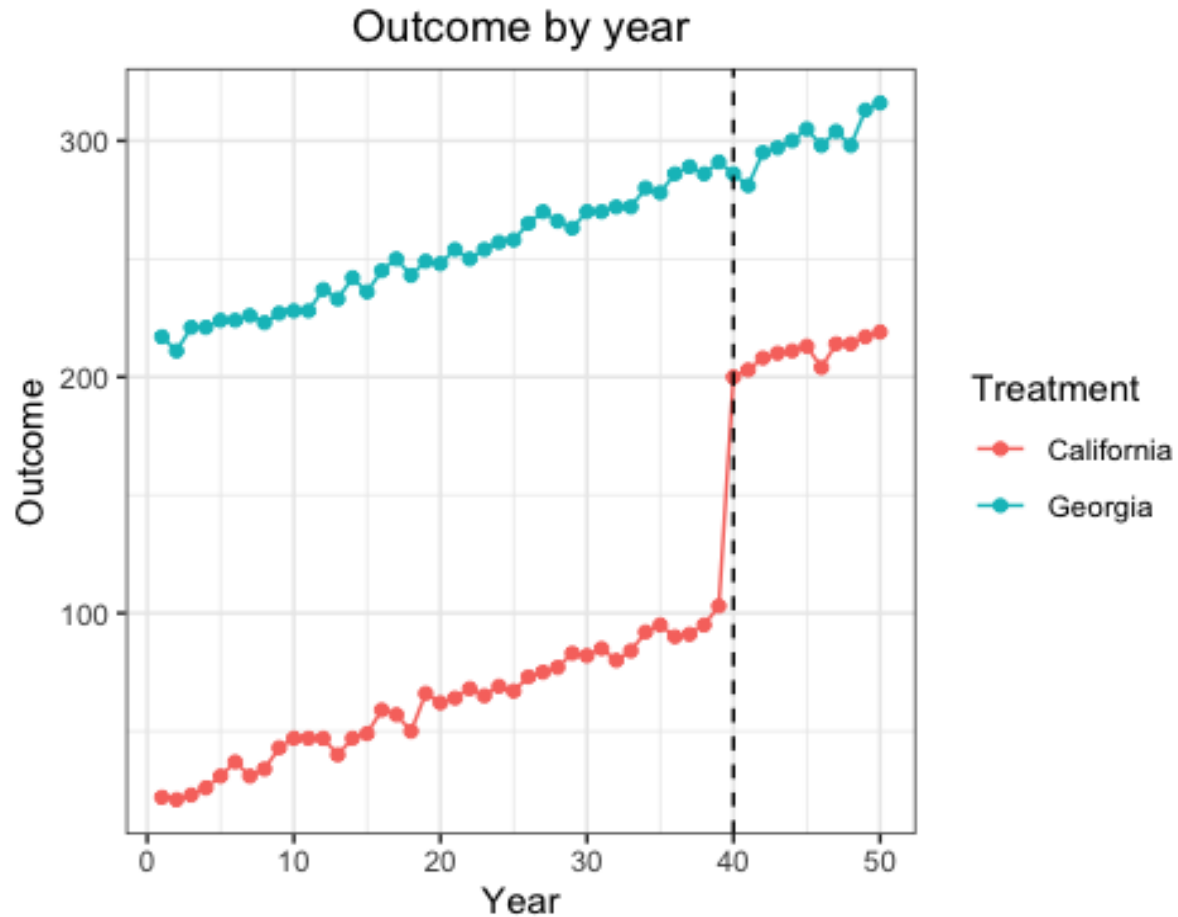 - Wildfires

2 main types of natural experiments
- **<u>Timing of the intervention</u>**
- Eligibility to a specific policy (e.g. age for vaccination) and other IV strategies

DIFFERENCE-IN-DIFFERENCES

The difference-in-differences idea

Nianogo, Benmarhnia and O'Neill [In preparation]

By Bret Zeldow and Laura Hatfield

DIFFERENCE-IN-DIFFERENCES
THE STANDARD APPROACH

- To estimate a causal effect

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

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

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

- Simple Regression modeling approach (with 2 groups)
  - $E[Y \mid a, t, did] = \beta_0 + \beta_1 a + \beta_2 t + \beta_3 did$
  - Where:
    - a represents the group with the policy
    - t epresents 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	$B'_0 + \beta_2$	$B'_0 + \beta'_1 + \beta'_2 + \beta_3$

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

0 0
1 0
0 1
1 1

Like a double fixed effect

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

MANY APPLICATIONS

Difference-in-Differences Analysis of the Association Between State Same-Sex Marriage Policies and Adolescent Suicide Attempts

Removing user fees for facility-based delivery services: a difference-in-differences evaluation from ten sub-Saharan African countries

Britt McKinnon,[1]* Sam Harper,[1] Jay S Kaufman[1] and Yves Bergevin[2]

Texting Bans and Fatal Accidents on Roadways: Do They Work? Or Do Drivers Just React to Announcements of Bans?[†]

By Rahi Abouk and Scott Adams*

Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study

Katherine B Howe,[1,2] Christian Suharlim,[3] Peter Ueda,[4,5] Daniel Howe, Ichiro Kawachi,[2] Eric B Rimm[1,6,7]

Europeanization process impacts the patterns of alcohol consumption in the Western Balkans

Eni Tresa[1,2], Tarik Benmarhnia[3], Timo Clemens[1], Genc Burazeri[1,4], Katarzyna Czabanowska[1,5]

A Difference-in-Differences Approach to Assess the Effect of a Heat Action Plan on Heat-Related Mortality, and Differences in Effectiveness According to Sex, Age, and Socioeconomic Status (Montreal, Quebec)

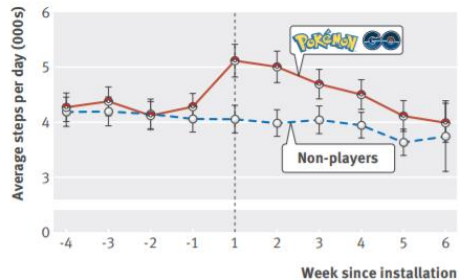Tarik Benmarhnia,[1] Zinzi Bailey,[1] David Kaiser,[2] Nathalie Auger,[3] Nicholas King,[4,5] and Jay S. Kaufman[1,5]

Fig 1 | Average number of daily steps and 95% confidence intervals by week before and after installation of Pokémon GO (median 8 July 2016)

DID COUPLED WITH PROPENSITY SCORE METHODS

- When multiple control groups are available, it is possible to use available information on time-varying and time-fixed confounders

- We can use propensity score matching and IPTW for example

- By doing so, we aim at identifying similar observations at each time point in the control groups

Using propensity scores in difference-in-differences models to estimate the effects of a policy change

Elizabeth A. Stuart · Haiden A. Huskamp · Kenneth Duckworth ·
Jeffrey Simmons · Zirui Song · Michael E. Chernew · Colleen L. Barry

LETTER

Quantifying the impact of changing the threshold of New York City heat emergency plan in reducing heat-related illnesses

Tarik Benmarhnia[1,3] [iD], Lara Schwarz[1], Amruta Nori-Sarma[2] [iD] and Michelle L Bell[2]
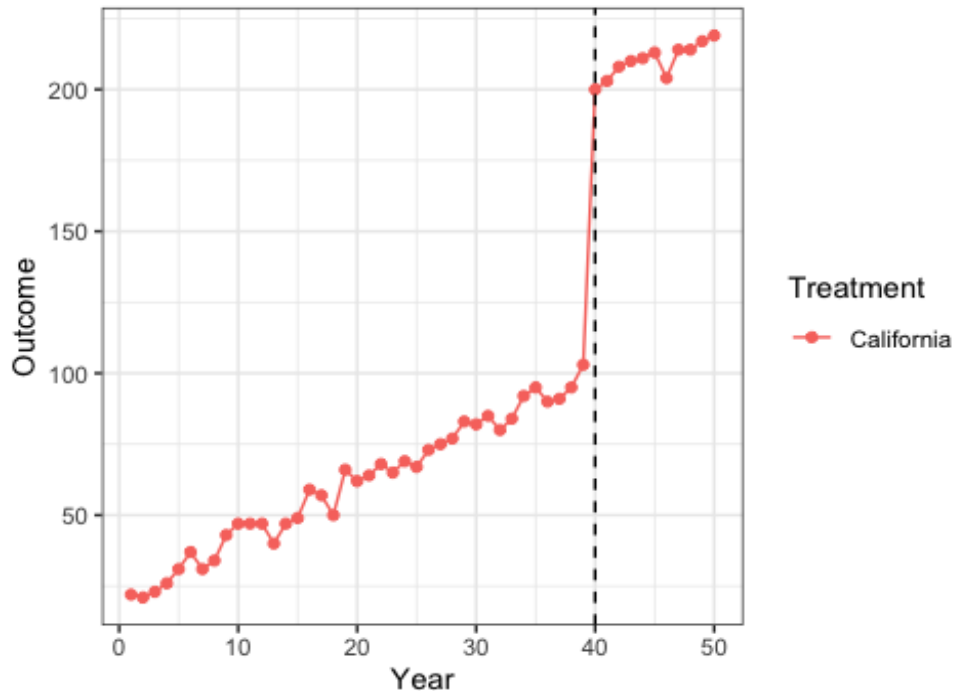
Evaluating the potential public health impacts of the Toronto cold weather program[*]

Tarik Benmarhnia[a,*], Xu Zhao[b], John Wang[b], Melissa Macdonald[c], Hong Chen[b,d,e]

INTERRUPTED TIME SERIES: DID <u>WITHOUT</u> CONTROL GROUPS

Outcome by year

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
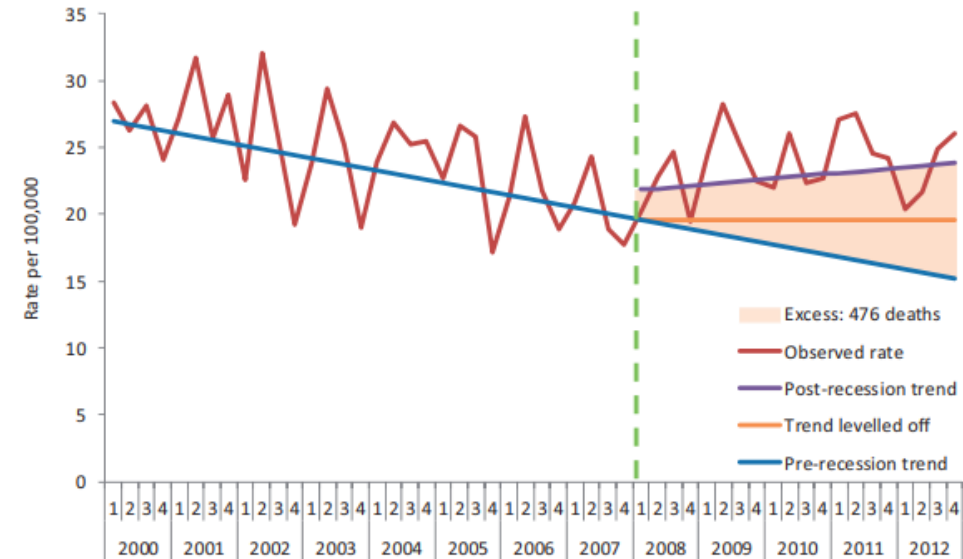2. Predicting Y in the post-treatment period using the model developed in stage 1 and compare with observed outcomes

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

ITS EXAMPLES

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

Impact of the economic recession and subsequent austerity on suicide and self-harm in Ireland: An interrupted time series analysis

Paul Corcoran,[1,2]* Eve Griffin,[1] Ella Arensman,[1,2] Anthony P Fitzgerald,[2] and Ivan J Perry[2]

CONTROLLED ITS OR DID?

- When using one (or more) control group.s, there is no fundamental distinction between Controlled ITS and DID models

- They yield the same counterfactuals and identify the same treatment effects.

- The terminology simply varies across disciplines

Fry, C. E., & Hatfield, L. A. Birds of a feather flock together: Comparing controlled pre–post designs. *Health Services Research*. https://onlinelibrary.wiley.com/doi/epdf/10.1111/1475-6773.13697
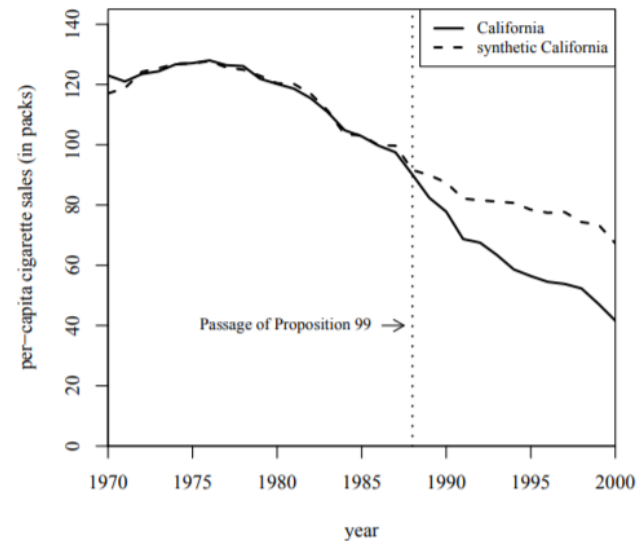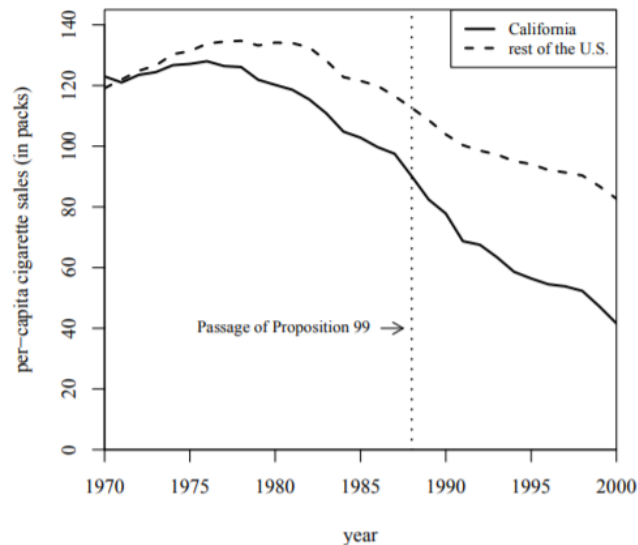
SYNTHETIC CONTROL METHODS

THE INTUITION

- When using a DID, it is sometimes difficult to establish whether the parallel trends assumption is met and whether the control group is a sufficiently accurate representation of what would have happened in the treated area without the intervention

- Synthetic control methodology (SCM) allows the construction of a counterfactual by selecting a **weighted average** of the outcome variable from a group of units similar to the treated unit

- The intuition behind this method using the original paper by Abadie et al. 2010

HOW DOES IT WORK?

- The synthetic control is based on the vector of weights **W** that minimizes the imbalance between the treated unit and a weighted average of the controls across a set of variables **X** (e.g. pre-intervention outcomes and/or covariates), $(X_1 - X_0 W)' V (X_1 - X_0 W)$

  - X_1 and X_0 contain the pre-treatment outcomes and covariates for the treated unit and control units respectively, and **V** captures the relative importance of these variables as predictors of the outcome of interest.

- In this setting, we assume the weights W to be positive and summing to 1 to avoid extrapolations issues (Abadie et al 2010). More recent approaches relaxed this assumption

- The treatment effect for the treated unit ($i = 1$), τ_{1t}, can then be estimated by $(Y_{1t} - \hat{Y}_{1t}^0)$ for each post-intervention period separately, and these can be averaged over time to obtain an ATT over the post-intervention period

- What happens when we have multiple treated units?
  - Discussed later …

STEPS IN CONDUCTING A SYNTHETIC CONTROL STUDY

1. Ensure the theory behind the intervention is well understood. Develop or present a conceptual model to make the theory transparent.
   - To ensure areas that have also been exposed to a similar intervention are excluded from the pool of potential controls

2. Identify potential control units that are plausibly eligible

3. Develop the synthetic control.
   - An optimization procedure using the outcome variables from the potential control areas to select the best weighting of units from the donor pool to create a synthetic control

4. Run outcome analysis and present results

5. Run robustness checks

KEY ASSUMPTIONS

Synthetic control methodology as a tool for evaluating population-level health interventions

Janet Bouttell,[1] Peter Craig,[2] James Lewsey,[1] Mark Robinson,[3] Frank Popham[2]

Table 2 Key assumptions of synthetic control methodology

Assumption	Assessment
1. Treated units and potential control units in the donor pool are similar.	Similar levels in variables known to influence outcome variable (see box 1 for objective and subjective elements of this assessment).
2. There is no contamination – spillover of effects of intervention into potential control units.	Based on background knowledge of researchers.
3. No external shocks in potential control units.	Based on background knowledge of researchers informed by review of trends in outcome variable.

EXAMPLES

A New Tool for Case Studies in Epidemiology—the Synthetic Control Method

David H. Rehkopf,[a] and Sanjay Basu[b]

Original Contribution

Health Behaviors, Mental Health, and Health Care Utilization Among Single Mothers After Welfare Reforms in the 1990s

Do medical marijuana laws reduce addictions and deaths related to pain killers?☆

David Powell [a],*, Rosalie Liccardo Pacula [a,b], Mireille Jacobson [b,c]

[a] RAND, Santa Monica, United States
[b] NBER, Cambridge, MA, United States
[c] University of California, Irvine, United States

Sanjay Basu*, David H. Rehkopf, Arjumand Siddiqi, M. Maria Glymour, and Ichiro Kawachi

Effects of changes in permit-to-purchase handgun laws in Connecticut and Missouri on suicide rates

CrossMark

Cassandra K. Crifasi *, John Speed Meyers, Jon S. Vernick, Daniel W. Webster

Johns Hopkins Center for Gun Policy and Research, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

OVERVIEW:
COMPARING DIFFERENT APPROACHES TO EVALUATE NATURAL EXPERIMENTS BASED ON THE TIMING OF THE INTERVENTION

- Uncontrolled before and after designs
  - Single or multiple measurements
- Interrupted time series design

No Control Group

- Controlled Before and after with single measurements
- Difference-in-Differences

With one Control Group

- Augmented DID approaches
  - With propensity score methods
- Synthetic Control Methods

With >1 Control Groups

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
    - Covered today
  - Treated units received the intervention at different times
    - Not covered today

- 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 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)

EXTENSIONS OF TRADITIONAL SC: GENERALIZED SYNTHETIC CONTROL (GSC)

Xu et al. 2017 proposed generalized synthetic control (**GSC**): estimates the average treatment effect on the treated using time-series cross sectional data

Improves efficiency and interpretability from SC, and can be used with multiple treated units and time varying confounders

GSC overcomes limitations in SC:

- Only unbiased when weights yield exact balance on lagged outcomes (and unidirectional weights)
- Only handles one treated unit at a time
- Non-parametric: no formal measures of variance

ADVANTAGES OF GSC

Traditional regression methods and SC are not well suited to estimate health impacts of extreme weather events (drought, heat wave, wildfire)

- Time varying confounding: (e.g. temperature, air pollution, other weather events)
- Widespread exposure - multiple exposed units

ESTIMATION STEPS

Estimation can be thought of as an out of sample prediction:

1) Estimate an interactive fixed effects model using only control group data to obtain latent factors

2) Project factor loadings onto each of these factors for the treated unit

3) Treated counterfactuals are estimated based on the estimated factors and factor loadings

$$Y_{it} = \delta_{it} D_{it} + x'_{it}\beta + \lambda'_i f_t + e_{it}$$

- D_{it} treatment indicator
- x_{it} observed covariates
- f_t latent factors
- λ_i factor loadings

$$Y^0_{it} = x'_{it}\beta + \lambda'_i f_t + e_{it}$$

$$Y^1_{it} = \delta_{it} + x'_{it}\beta + \lambda'_i f_t + e_{it}$$

$$\widehat{\Delta}_{it} = Y^1_{it} - \widehat{Y}^0_{it}$$

CASE STUDY: 2007 SOUTHERN CALIFORNIA WILDFIRES ON RESPIRATORY HOSPITALIZATIONS [SHERIDAN ET AL.]

As the climate changes, wildfires are expected to increase in frequency, intensity and duration (especially in California)

In this case study we use:

- Satellite based smoke plume data and burn area data to classify wildfire exposure
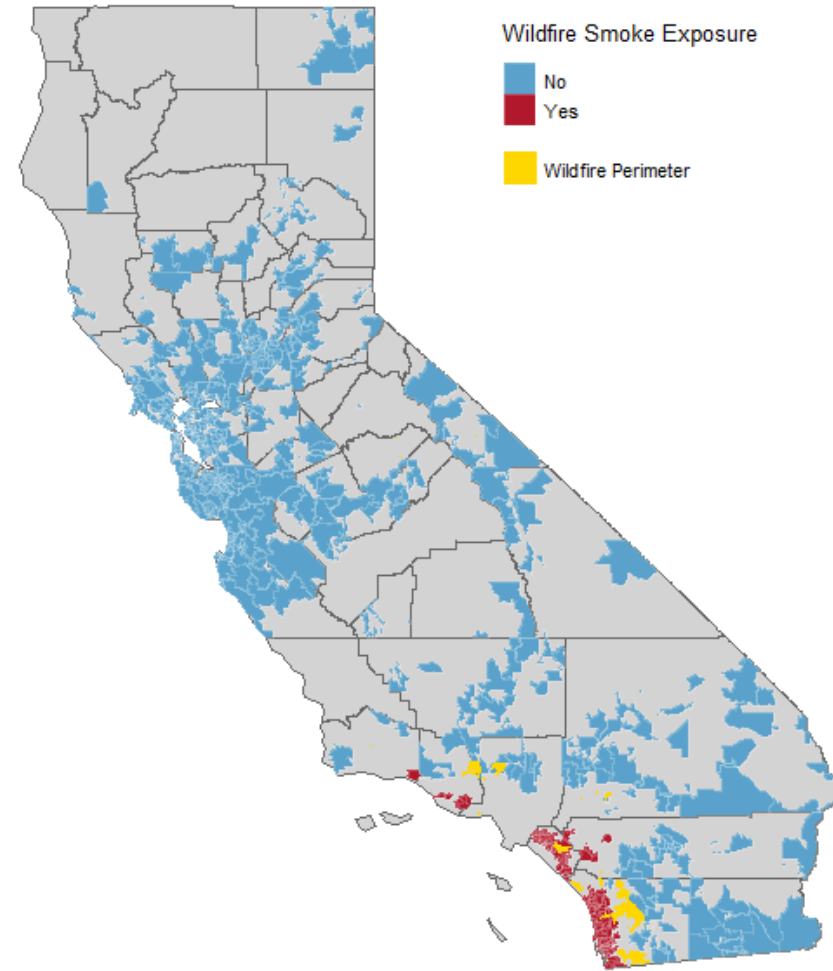- OSHPD respiratory hospitalization data by zip code for outcome

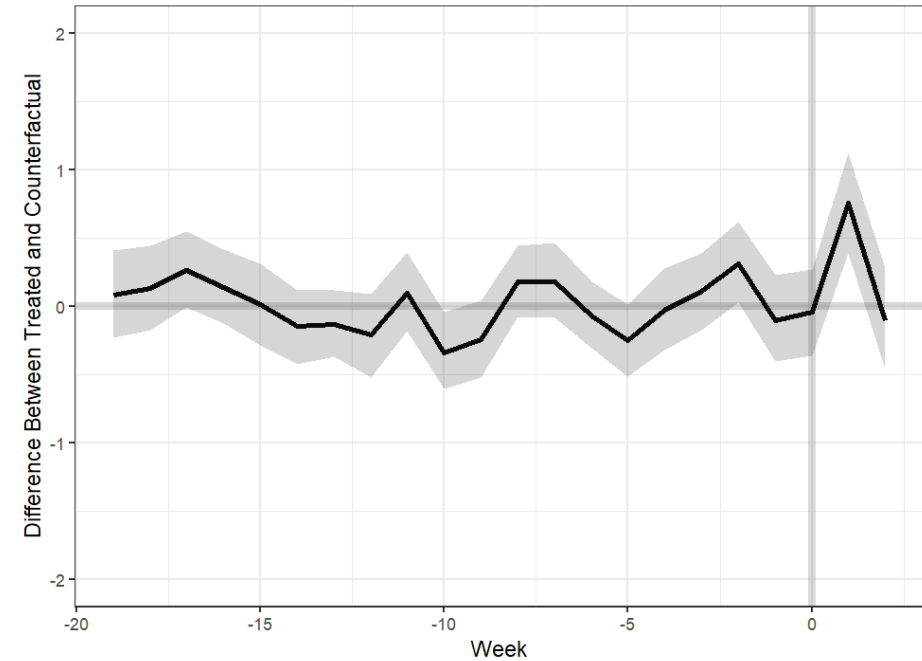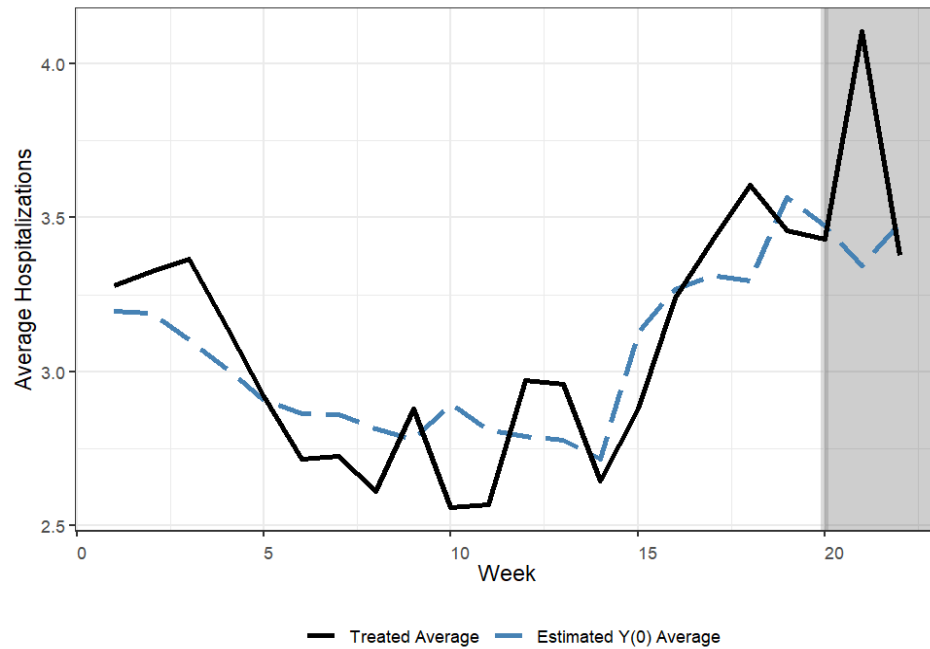# RESULTS

Out of 1779 zip code
tabulation areas in
California

685 were included:
- 130 exposed
- 555 unexposed

Wildfire Smoke Exposure
- No
- Yes

Wildfire Perimeter

Map of California zip codes exposed to wildfire smoke week of October 20th 2007.
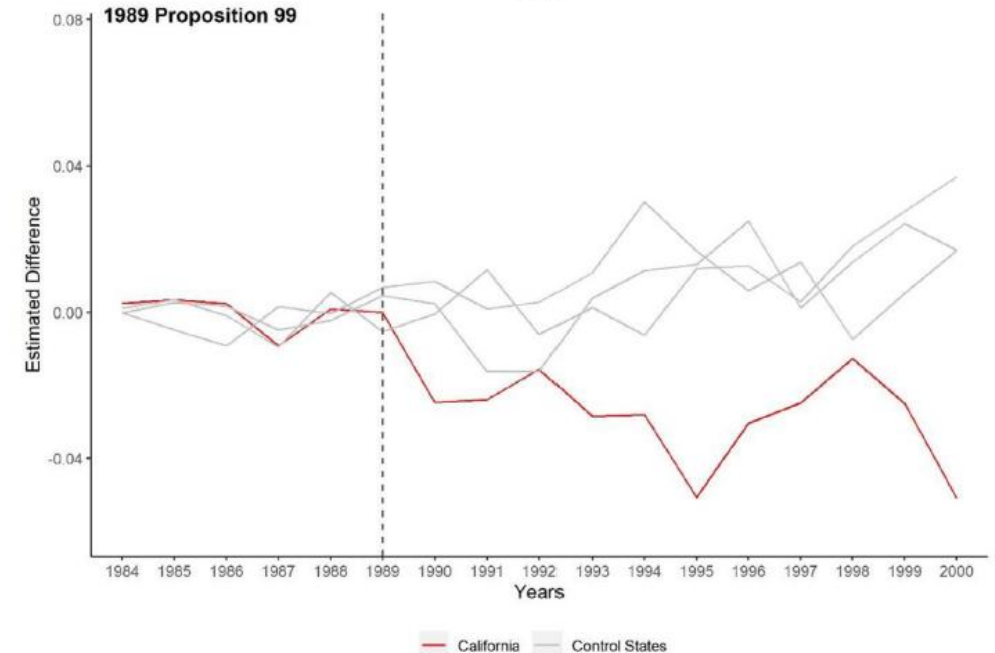
RESULTS

We found an 18% (95%CI: 10%-29%) average increase in respiratory hospitalizations as a result of the wildfire storm

THE IMPORTANCE OF FALSIFICATION TESTS

- <u>Permutation tests</u>

- Many assumptions cannot be checked empirically
  - It is therefore important to design a set of falsification/placebo tests to improve the inference of interest

- Negative control approaches
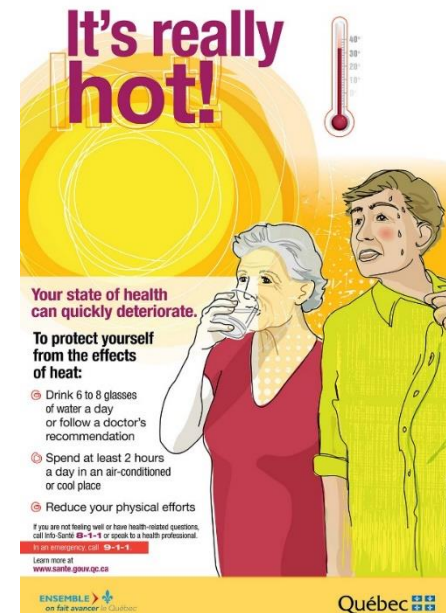 - Lipsitch M et al. (2012): Negative controls: a tool for detecting confounding and bias in observational studies.

Sheridan et al. "Evaluating the impact of the California 1995 smoke-free workplace law on population smoking prevalence using a synthetic control method." Preventive medicine reports 19 (2020): 101164.

A Difference-in-Differences Approach to Assess the Effect of a Heat Action Plan on Heat-Related Mortality, and Differences in Effectiveness According to Sex, Age, and Socioeconomic Status (Montreal, Quebec)

Tarik Benmarhnia,[1] Zinzi Bailey,[1] David Kaiser,[2] Nathalie Auger,[3] Nicholas King,[4,5] and Jay S. Kaufman[1,5]

- The Montreal heat warning system
  - Implemented in 2004
  - 'active watch' alert level, when daily max temperatures **exceed 30°C**
  - Focus on vulnerable populations
    - Age
    - SES
    - Gender as a placebo

THE ANALYTICAL APPROACH:

- Assigning days (the unit of analysis) to two groups: an "intervention" group that meets an eligibility criterion and a "non-intervention" group that does not.

- the "intervention" group: if 'active watch' alert level is present (heat wave days)

- The non-intervention group: non heat wave days

- The **counterfactual quantity** being estimated is:
  - The difference in the daily number of deaths between intervention (heat wave) and non-intervention (non-heat wave) days in the post-2004 period, had the HAP not been implemented

SENSITIVITY ANALYSES

Table 3. Sensitivity analyses for the estimated effects of the heat action plan program.

Sensitivity analyses	DID estimate	95% CI	p-Value[a]
Arbitrary programs			
Program implemented in 2000[b]	0.94	(−2.08, 3.96)	0.54
Program implemented in 2002[c]	0.42	(−3.62, 2.77)	0.80
Other hot days definitions			
When maximum temperature is above 28°C	0.58	(−1.77, 2.93)	0.63
When maximum temperature is above 32°C	2.79	(−2.65, 8.23)	0.32
Cumulative heat[d]	4.87	(0.67, 8.20)	0.03
Accounting for displacement ratio[e]	1.87	(0.29, 3.47)	0.02
Restriction to non-eligible days above 25°C	2.23	(−0.80, 5.27)	0.15

[a]p-Values are obtained from a Wald test on the interaction term (i.e., DID estimate).
[b]Using mortality and temperature data for periods 1996–1999 vs. 2000–2003.
[c]Using mortality and temperature data for periods 1998–2001 vs. 2002–2005.
[d]Considering a cumulative heat effect up to 5 consecutive hot days (lag 0–5).
[e]The displacement ratio (Saha et al. 2014) was 0.65.

TOPICS NOT DISCUSSED TODAY

- 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

- Multi-stage interventions

- The Scale of interest
  - Important to think about the scale of the interaction term
    - When Is Parallel Trends Sensitive to Functional Form? (Roth and Sant'Anna, 2021)

- Lagged/delayed effects
  - Many available tools

DESIGNING A STUDY CAPITALIZING ON THE TIMING OF AN INTERVENTION

When designing a study in which the timing of an intervention is capitalized on, we recommend to undertake the three following steps

1. Define a clear research question, ideally with a well-defined intervention with specific implementation time and targeted population;
   - ✓ The target trial framework which considers the design of an ideal randomized experiment, can be helpful in framing the study and the target causal estimand of interest.

2. Propose an identification strategy about how we can exploit the setting of the intervention implementation of interest to estimate the targeted causal quantity of interest (estimand) and identify possible falsifications tests

3. Propose an estimation strategy which often relates to the statistical or modelling approach (i.e. estimator) that we will use to obtain a given estimate (e.g. regression coefficient) to quantitatively represent the targeted estimand of interest (i.e. ATT).

STEP BY STEP HYPOTHETICAL EXAMPLE

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed
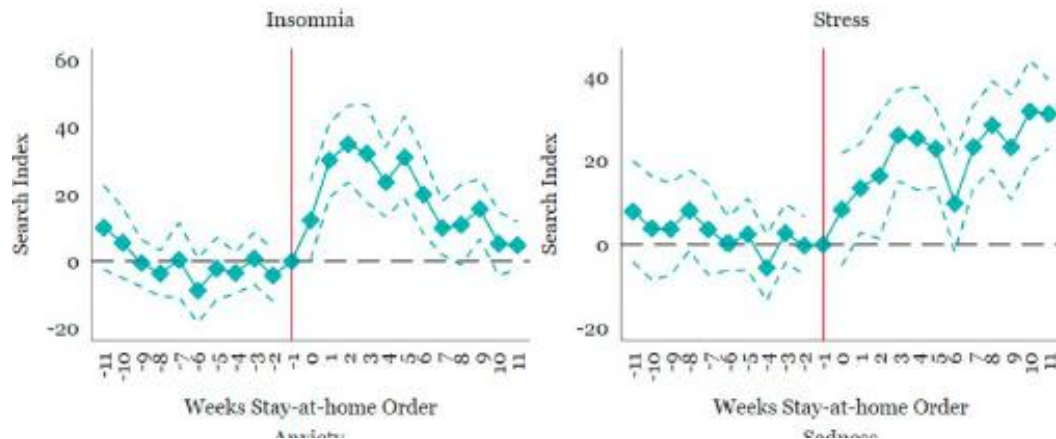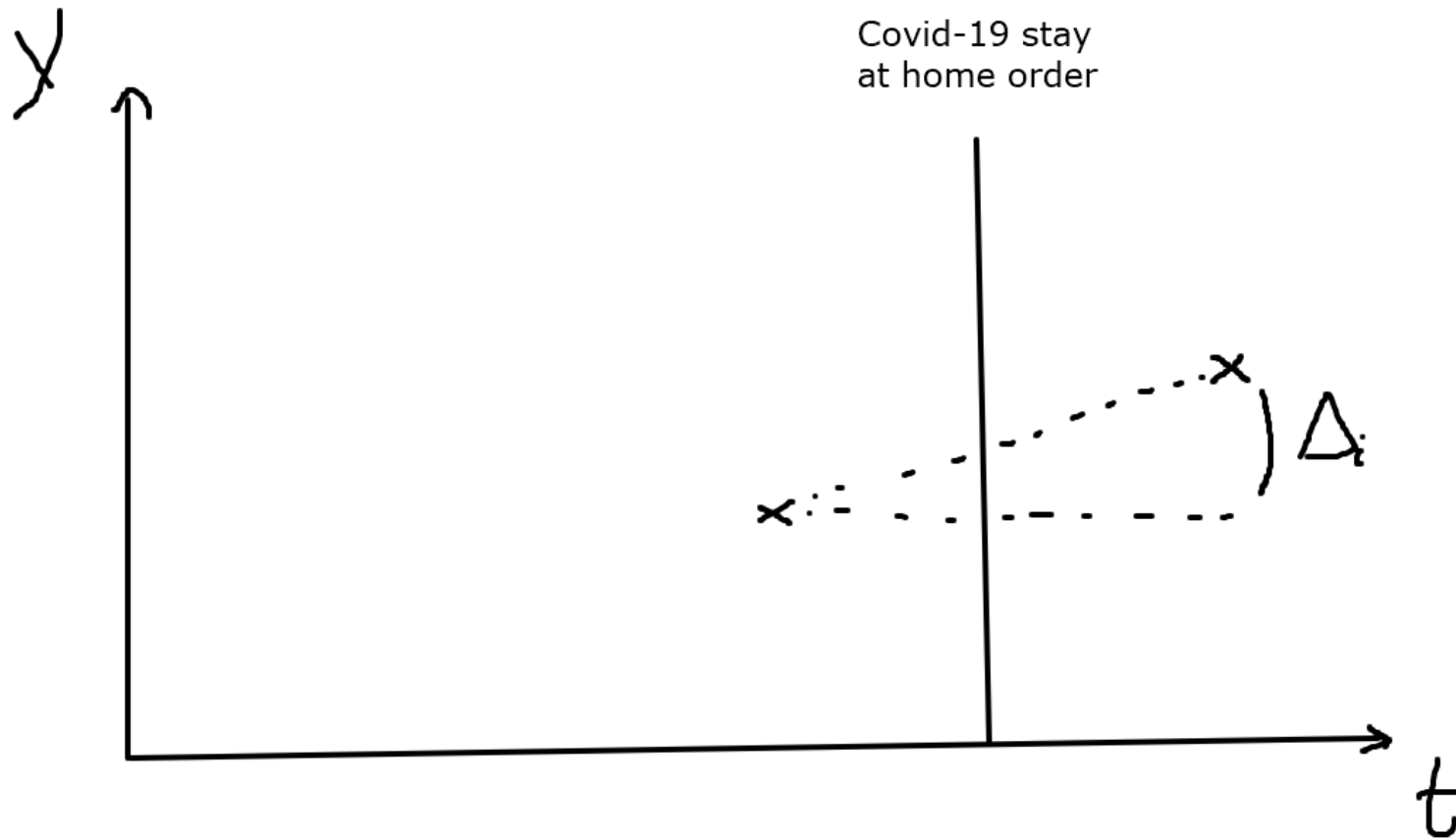
COVID-19 blues: Lockdowns and mental health-related google searches in Latin America
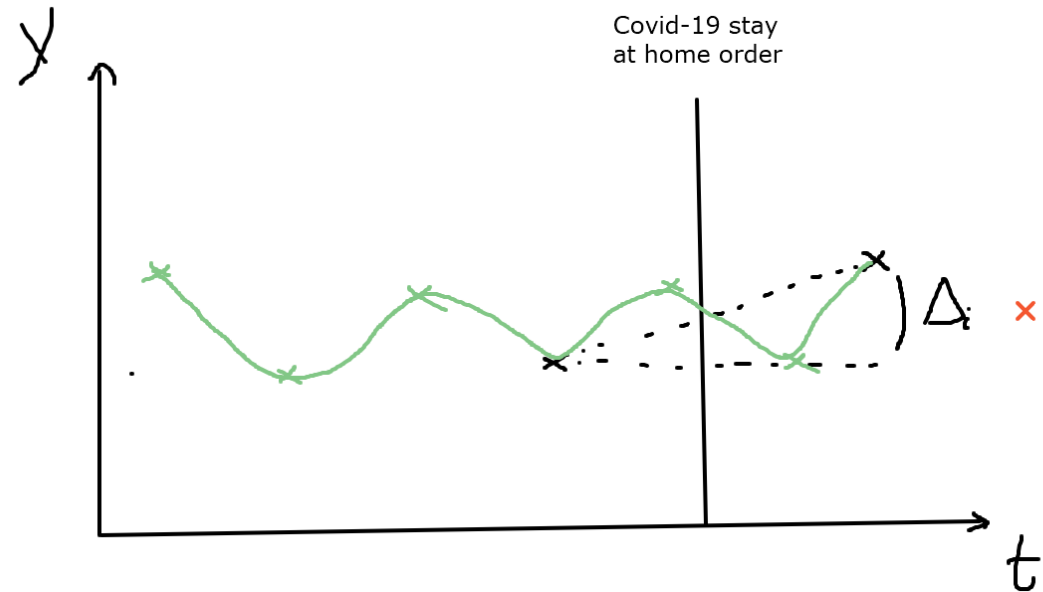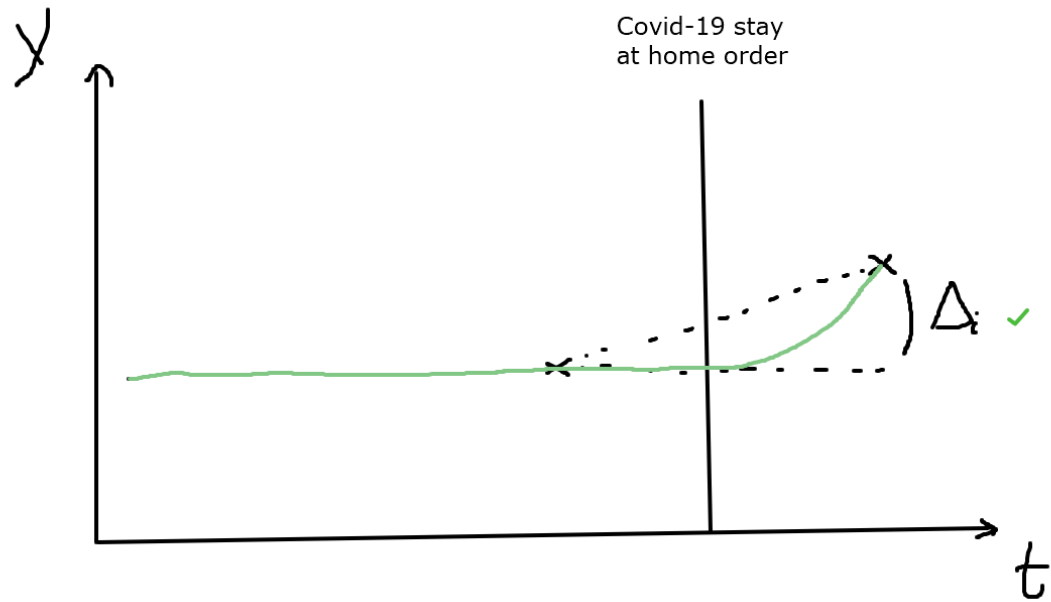
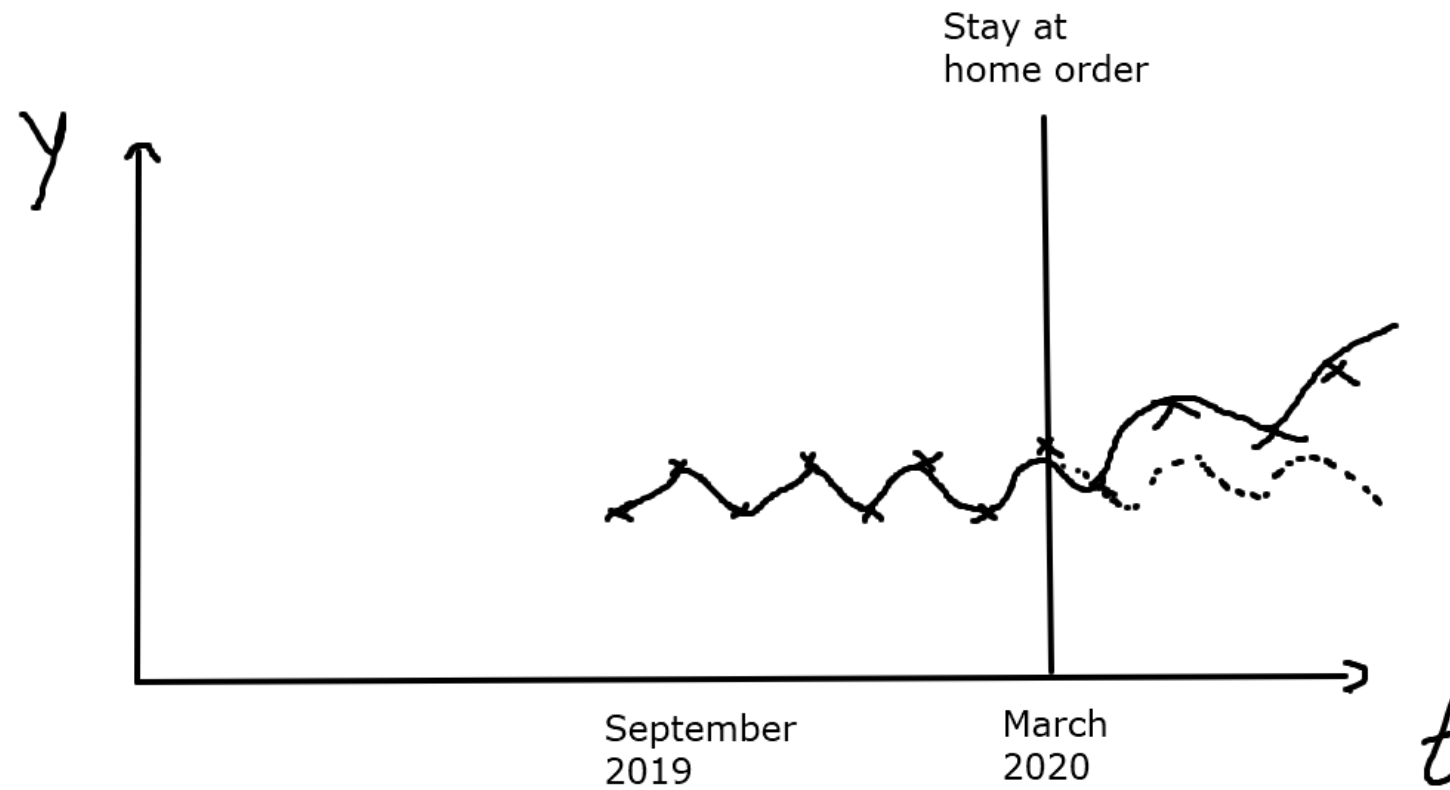Adan Silverio-Murillo [a], Lauren Hoehn-Velasco [b], Abel Rodriguez Tirado [a], Jose Roberto Balmori de la Miyar [c,*]
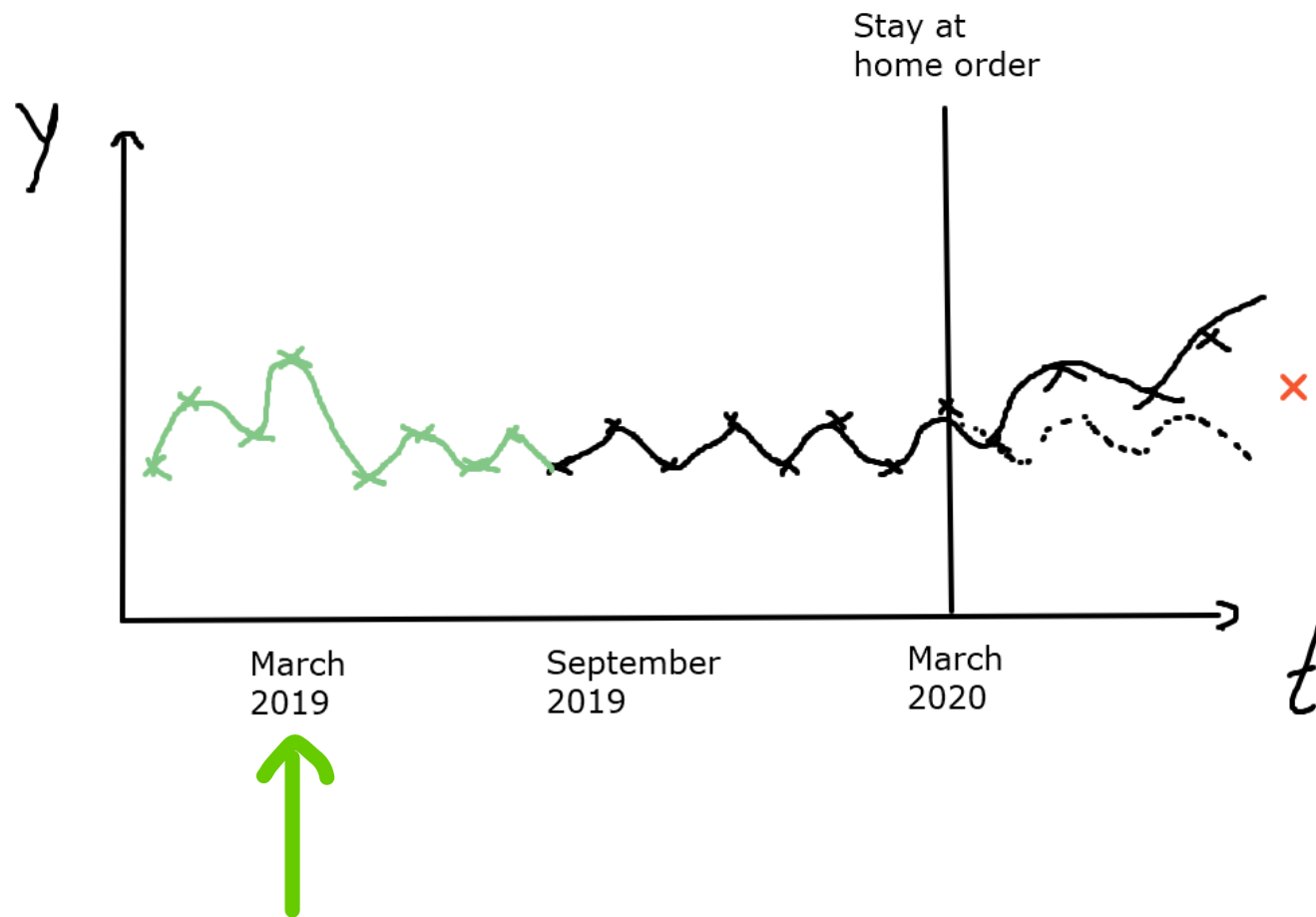
SIMPLE BEFORE AND AFTER DESIGN

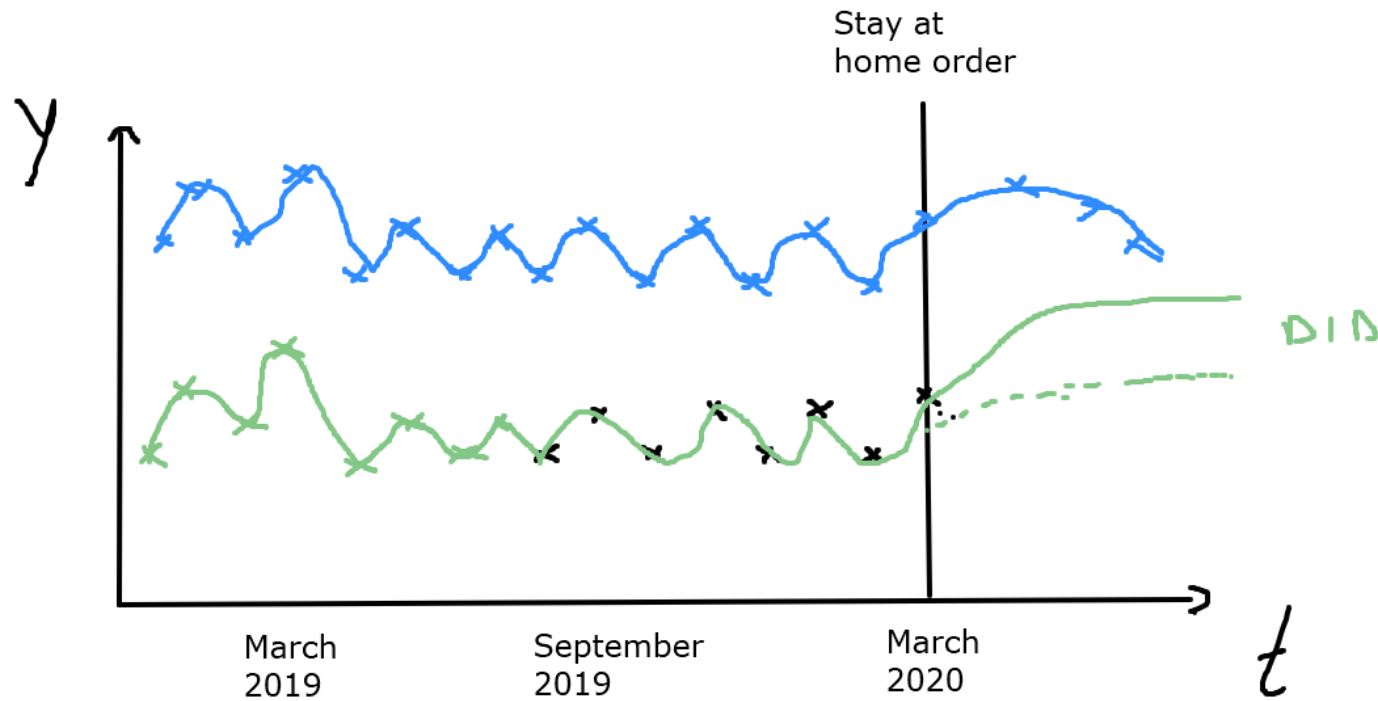# WHAT COULD GO WRONG

USING AN ITS APPROACH
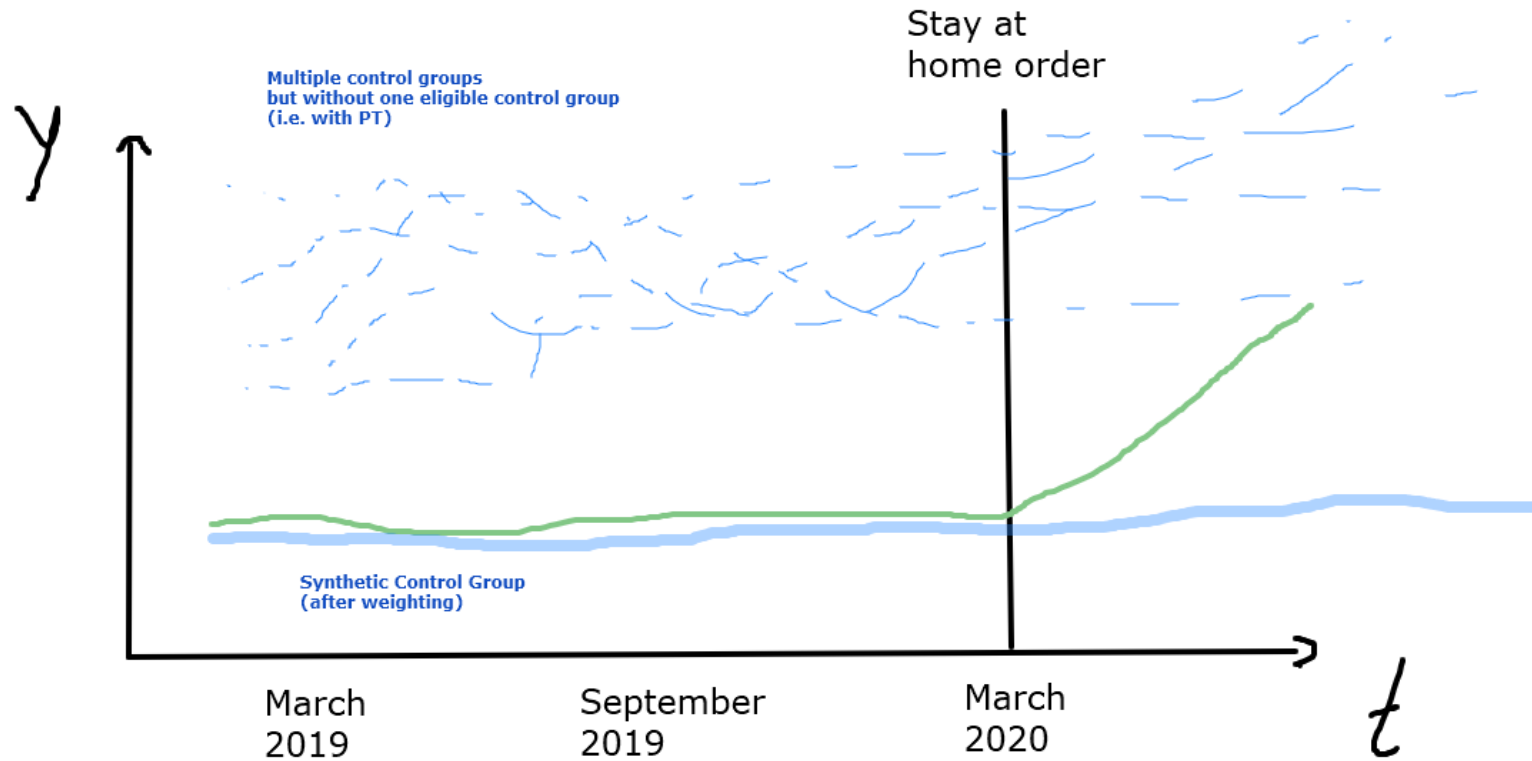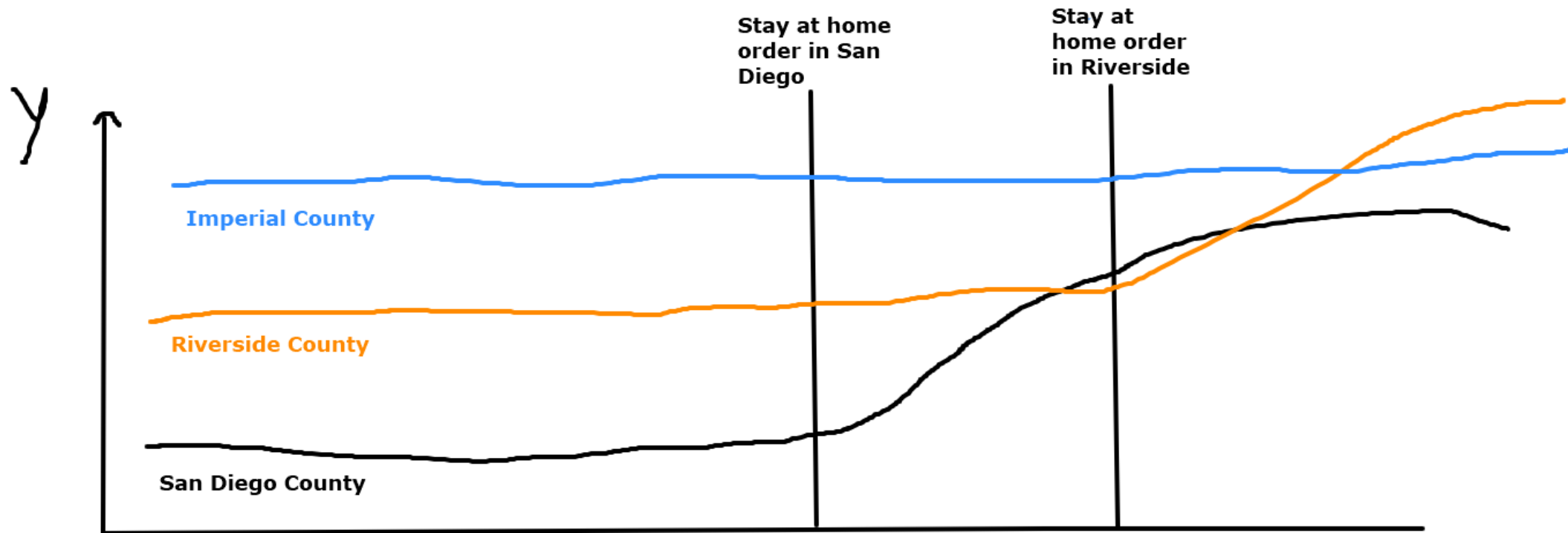
WHAT COULD GO WRONG

SOLUTION: DID OR CITS
IF YOU HAVE AN ELIGIBLE CONTROL GROUP (WITH PT)

WITH MULTIPLE CONTROL GROUPS

WITH VARIATION IN THE TIMING OF THE INTERVENTION

REVISITING THE WORKSHOP OBJECTIVES

1. Understand the conceptual foundations of quasi-experimental methods based on the timing of a natural experiment

2. Understanding how Difference-in-Differences (DID), Interrupted Time Series (ITS) and Synthetic Control Methods (SCM) work and their specific identification assumptions.

3. Being able to implement (in R) DID, ITS and SCM in different settings, interpret estimands of interest and check/visualize possible assumptions' violations
   ✓ **Coming next**

REFERENCES

- Craig, P., Katikireddi, S. V., Leyland, A., & Popham, F. (2017). Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annual review of public health, 38, 39-56.

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American statistical Association, 105(490), 493-505.

- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. Political Analysis, 25(1), 57-76.

- Acemoglu, D., Johnson, S., Kermani, A., Kwak, J. and Mitton, T., 2016. The value of connections in turbulent times: Evidence from the United States. Journal of Financial Economics, 121(2), pp.368-391.

- Dube A, Zipperer B. Pooled synthetic control estimates for recurring treatment: an application to minimum wage studies, University of Massachusetts. Amherst Working Paper.2013

- Kreif, N., Grieve, R., Hangartner, D., Turner, A.J., Nikolova, S. and Sutton, M., 2016. Examination of the synthetic control method for evaluating health policies with multiple treated units. Health economics, 25(12), pp.1514-1528.

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics.

- Callaway, B., & Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. Journal of Econometrics.

- Roth, J., & Sant'Anna, P. H. (2020). When Is Parallel Trends Sensitive to Functional Form?. arXiv preprint arXiv:2010.04814.

- Lipsitch, M., Tchetgen, E. T., & Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology (Cambridge, Mass.), 21(3), 383.

THANK YOU

NIAROCH@UCLA.EDU
TBENMARHNIA@UCSD.EDU