Causal Inference
Designs

José R. Zubizarreta
Harvard University

09/04/2023
CUSO Doctoral School in Statistics and Applied Probability
Saignelégier, Switzerland

Outline

Overview

- Causal inference
  - Which treatments work?
  - Form whom?
  - When?
  - And why?
- Interventions
  - Point exposures
  - Time-varying
- Designs
  - Experimental
  - Observational
- Strategies
  - Randomization
  - Observation, assumptions
    - E.g., instruments
- Core methods
  - Matching
  - Regression
  - Weighting
- Sensitivity analyses
- Evidence integration
- Issues throughout
  - Missingness
  - Mismeasurement
  - Fairness...
- Perspectives
  - Statistics, biostatistics
  - Economics, political science
  - Computer science

Outline

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data
  - This, regardless of the method used (either sub-classification, matching, g-computation...)

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data
  - This, regardless of the method used (either sub-classification, matching, g-computation...)
- Now, imagine for a moment there was an alternative that allowed us to make causal inference even if we do not adjust for all relevant covariates

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data
  - This, regardless of the method used (either sub-classification, matching, g-computation...)
- Now, imagine for a moment there was an alternative that allowed us to make causal inference even if we do not adjust for all relevant covariates
- This would be a dream, right?

Immaculate results?

- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data
  - This, regardless of the method used (either sub-classification, matching, g-computation...)
- Now, imagine for a moment there was an alternative that allowed us to make causal inference even if we do not adjust for all relevant covariates
- This would be a dream, right?
- Well, this is the promise of an instrumental variable!

Immaculate results?

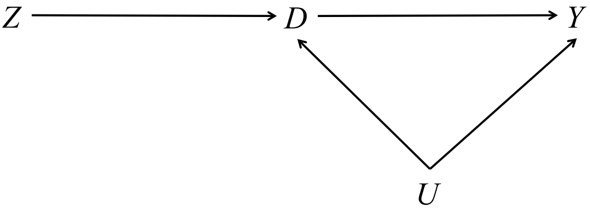
- Can you guarantee that the results of your observational study are not affected by an unobserved covariate?
- No matter how pristine is the design of an observational study, the assumption of "no unmeasured confounders" is typically key to making causal inferences from non-experimental data
  - This, regardless of the method used (either sub-classification, matching, g-computation...)
- Now, imagine for a moment there was an alternative that allowed us to make causal inference even if we do not adjust for all relevant covariates
- This would be a dream, right?
- Well, this is the promise of an instrumental variable!
- See Hernán and Robins (2006) for a discussion

Overview of instrumental variables

- What is an instrument?
  - A haphazard push to receive treatment which affects the outcome only through the treatment

Overview of instrumental variables

- What is an instrument?
  - A haphazard push to receive treatment which affects the outcome only through the treatment
- Main assumptions
  - (R) The push is essentially random after adjusting for observed covariates
  - (E) The push affects the outcome only through the treatment (exclusion restriction)



Examples of instrumental variables in healthcare studies

- Sources of instrumental variables in health studies include (Baiocchi et al. 2014)
  - Randomized encouragement trials
  - Distance to specialty care provider
  - Preferences in medical practice
  - Calendar time
  - Genes
  - Timing of admission
  - Insurance plan
  - Congestion

The randomized encouragement design

- Perhaps the most basic example of an instrumental variable appears in Holland's (1988) randomized encouragement design
  - Here, subjects are randomized to one of two groups

The randomized encouragement design

- Perhaps the most basic example of an instrumental variable appears in Holland's (1988) randomized encouragement design
  - Here, subjects are randomized to one of two groups
  - Members of one group are encouraged, say, to quit smoking

The randomized encouragement design

- Perhaps the most basic example of an instrumental variable appears in Holland's (1988) randomized encouragement design
  - Here, subjects are randomized to one of two groups
  - Members of one group are encouraged, say, to quit smoking
  - The outcome (e.g. an evaluation of lung tissue) might respond to a reduction in cigarettes consumed but not to the encouragement per se

The randomized encouragement design

- Perhaps the most basic example of an instrumental variable appears in Holland's (1988) randomized encouragement design
  - Here, subjects are randomized to one of two groups
  - Members of one group are encouraged, say, to quit smoking
  - The outcome (e.g. an evaluation of lung tissue) might respond to a reduction in cigarettes consumed but not to the encouragement per se
- The Wald estimator attributes the entire difference in outcomes between the randomized encouraged and unencouraged groups to the greater change in behavior in the encouraged group

$$\hat{\tau}_{\text{Wald}} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

Compliance types

		Encouraged (Z=1)	
		Treated ($D(1)$=1)	Not treated ($D(1)$=0)
Discouraged (Z=0)	Treated ($D(0)$=1)	**Always-taker** ($D(0)=D(1)=1$)	**Defier** ($D(0)>D(1)$)
	Not treated ($D(0)$=0)	**Complier** ($D(0)<D(1)$)	**Never-taker** ($D(0)=D(1)=0$)

Estimation

- Conventionally, the Complier Average Causal Effect (CACE) is estimated by two-stage least squares (2SLS) regression
- Problem: weak instruments
  - Confidence intervals derived from 2SLS do not have adequate coverage
    - Fixable with appropriate inferential methods (Imbens and Rosenbaum 2005)
  - Estimates are sensitive to very small biases
    - This problem does not go away in large samples (Small and Rosenbaum 2008)

Questions?

- How to strengthen an instrument?
- How to target other estimands?
- How to characterize heterogeneity?
- How to handle instruments in longitudinal studies?

Outline

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates
- However, if there are unobserved covariates, then we need to look for other approaches

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates
- However, if there are unobserved covariates, then we need to look for other approaches
- One such approach is an instrumental variable

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates
- However, if there are unobserved covariates, then we need to look for other approaches
- One such approach is an instrumental variable
- A related approach is a discontinuous treatment assignment rule, whereby units with a value of an assignment, forcing, or running variable above a certain cutoff value are assignment to treatment, and otherwise are assigned to control

Discontinuity designs

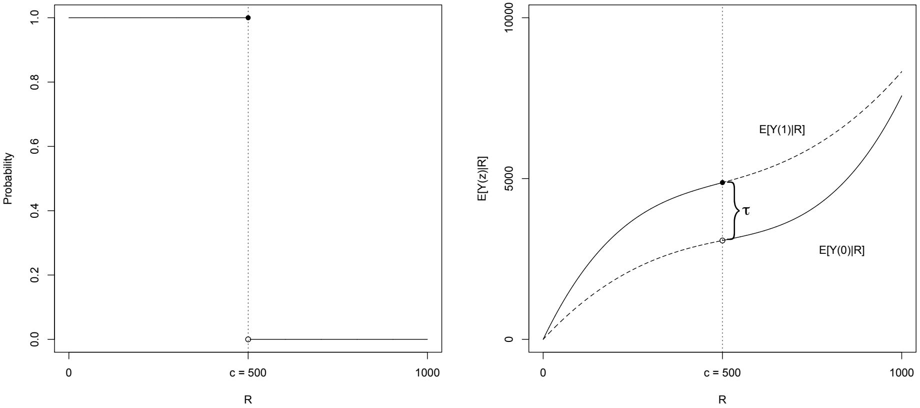
- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates
- However, if there are unobserved covariates, then we need to look for other approaches
- One such approach is an instrumental variable
- A related approach is a discontinuous treatment assignment rule, whereby units with a value of an assignment, forcing, or running variable above a certain cutoff value are assignment to treatment, and otherwise are assigned to control
- The idea is that in a neighborhood of the discontinuity, units treatment assignment is essentially random

Discontinuity designs

- Drawing causal inferences from non-experimental or observational data is a complex task
- If treatment assignment is governed only by observed covariates, then it suffices to adjust for differences in such covariates
- However, if there are unobserved covariates, then we need to look for other approaches
- One such approach is an instrumental variable
- A related approach is a discontinuous treatment assignment rule, whereby units with a value of an assignment, forcing, or running variable above a certain cutoff value are assignment to treatment, and otherwise are assigned to control
- The idea is that in a neighborhood of the discontinuity, units treatment assignment is essentially random
- Therefore by restricting the analyses to that neighborhood we can get an unbiased average treatment effect estimate

Idea of a discontinuity design

- In our labor training example, imagine subjects with pre-treatment income below 500 are assigned to the program
- Pre-treatment income is the running variable R, 500 is the cutoff c



The continuity-based framework

- RD designs are treated as an observational study with a non-probabilistic assignment mechanism

The continuity-based framework

- RD designs are treated as an observational study with a non-probabilistic assignment mechanism
- If the mean potential outcome functions are smooth functions of the running variable and other covariates (observed and unobserved ones), then the discontinuity in the conditional expectation function of the observed outcome given the running variable at the cutoff can be interpreted as the average treatment effect for units at the cutoff

The continuity-based framework

- RD designs are treated as an observational study with a non-probabilistic assignment mechanism
- If the mean potential outcome functions are smooth functions of the running variable and other covariates (observed and unobserved ones), then the discontinuity in the conditional expectation function of the observed outcome given the running variable at the cutoff can be interpreted as the average treatment effect for units at the cutoff
- See Thistlethwaite and Campbell (1960), Imbens and Lemieux (2008), Lee and Lemieux (2010), Imbens and Kalyanaraman (2012), Calonico et al. (2014), and Calonico et al. (2016)

The local randomization framework

- It assumes that the assignment mechanism is probabilistic, at least for units within a neighborhood around the cutoff value

The local randomization framework

- It assumes that the assignment mechanism is probabilistic, at least for units within a neighborhood around the cutoff value
- If the running variable is independent from the potential outcomes for the individuals within this neighborhood, then the RD design can essentially be treated as a randomized experiment for the units within the neighborhood

The local randomization framework

- It assumes that the assignment mechanism is probabilistic, at least for units within a neighborhood around the cutoff value
- If the running variable is independent from the potential outcomes for the individuals within this neighborhood, then the RD design can essentially be treated as a randomized experiment for the units within the neighborhood
- See Cattaneo et al. (2015), Li, Mattei, and Mealli (2015), and Mattei and Mealli (2016)

The local randomization framework

- It assumes that the assignment mechanism is probabilistic, at least for units within a neighborhood around the cutoff value
- If the running variable is independent from the potential outcomes for the individuals within this neighborhood, then the RD design can essentially be treated as a randomized experiment for the units within the neighborhood
- See Cattaneo et al. (2015), Li, Mattei, and Mealli (2015), and Mattei and Mealli (2016)
- In some cases, it is implausible that the mean potential outcome functions are constant functions of the running variables, even in a neighborhood of cutoffs

Questions?

- How to relax these assumptions?
- How to target other estimands?
- How to characterize heterogeneity?
- How to handle more complex treatment rules?

Impact of grade retention on juvenile crime

Click here for picture copyrights

Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
  - 7 is "Outstanding,"
  - 4 is "Sufficient,"
  - 1 stands for "Very Deficient"

Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
  - 7 is "Outstanding,"
  - 4 is "Sufficient,"
  - 1 stands for "Very Deficient"
- Grade retention is determined by two rules:

Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
  - 7 is "Outstanding,"
  - 4 is "Sufficient,"
  - 1 stands for "Very Deficient"
- Grade retention is determined by two rules:
  - Rule 1: a grade below 4 in one subject *and* an average grade across all subjects below 4.5

Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
  - 7 is "Outstanding,"
  - 4 is "Sufficient,"
  - 1 stands for "Very Deficient"
- Grade retention is determined by two rules:
  - Rule 1: a grade below 4 in one subject *and* an average grade across all subjects below 4.5
  - Rule 2: a grade below 4 in two subjects *and* an average grade across all subjects below 5

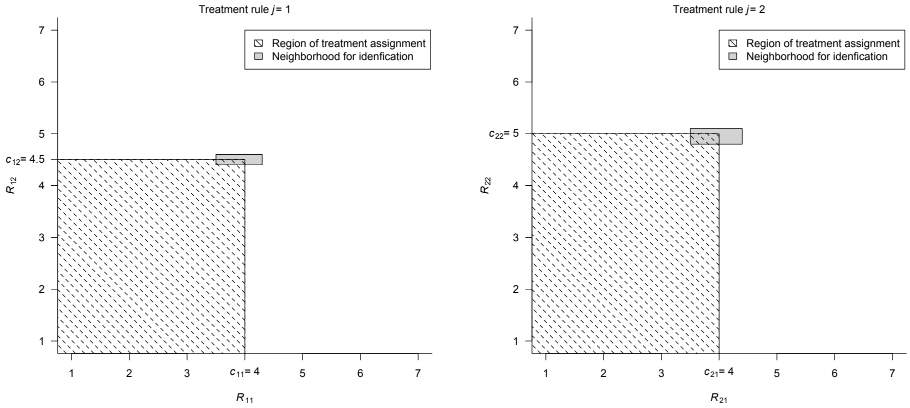
Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
  - 7 is "Outstanding,"
  - 4 is "Sufficient,"
  - 1 stands for "Very Deficient"
- Grade retention is determined by two rules:
  - Rule 1: a grade below 4 in one subject *and* an average grade across all subjects below 4.5
  - Rule 2: a grade below 4 in two subjects *and* an average grade across all subjects below 5
- This defines a "complex" discontinuity design, with multiple rules, several running variables, and many cutoffs that lead to the *same* treatment

Grade retention in Chile

- School grades vary between 1 and 7 by increments of 0.1:
    - 7 is "Outstanding,"
    - 4 is "Sufficient,"
    - 1 stands for "Very Deficient"
- Grade retention is determined by two rules:
    - Rule 1: a grade below 4 in one subject *and* an average grade across all subjects below 4.5
    - Rule 2: a grade below 4 in two subjects *and* an average grade across all subjects below 5
- This defines a "complex" discontinuity design, with multiple rules, several running variables, and many cutoffs that lead to the *same* treatment
- How to analyze these designs?

Visualizing the rules and selected neighborhoods



Assumptions for identification

Assumption 1 (A1). Local unconfoundedness of treatment assignment via the running variables

$$\mathcal{R}_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\} \mid \boldsymbol{X}_i, \mathcal{Q}_i \in \mathcal{N}$$

Assumptions for identification

Assumption 1 (A1). Local unconfoundedness of treatment assignment via the running variables

$$\mathcal{R}_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\} \mid \boldsymbol{X}_i, \mathcal{Q}_i \in \mathcal{N}$$

Assumption 2 (A2). Local positivity of treatment assignment

$$0 < \Pr(Z_i = 1 \mid \boldsymbol{X}_i, \mathcal{Q}_i \in \mathcal{N}) < 1$$

Assumptions for identification

Assumption 1 (A1). Local unconfoundedness of treatment assignment via the running variables

$$\mathcal{R}_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\} \mid \boldsymbol{X}_i, \mathcal{Q}_i \in \mathcal{N}$$

Assumption 2 (A2). Local positivity of treatment assignment

$$0 < \Pr(Z_i = 1 \mid \boldsymbol{X}_i, \mathcal{Q}_i \in \mathcal{N}) < 1$$

- For simple rules, similar assumptions have been considered by Battistin and Rettore (2008), Keele et al. (2015), Angrist and Rokkanen (2015), Mattei (2017), Branson and Mealli (2018)

Sensitivity analysis using a near-equivalence test

- In the absence of hidden bias, grade retention does not cause committing a juvenile crime

Table: NATE estimate

Outcome variable	Matched sample mean		$\widehat{\tau}_{\text{NATE}}$	$H_0 : \tau_{\text{NATE}} = 0$
	Treated	Control		One-sided p-value
Committing a crime	0.059	0.053	0.006	0.229

Sensitivity analysis using a near-equivalence test

- In the absence of hidden bias, grade retention does not cause committing a juvenile crime

Table: NATE estimate

Outcome variable	Matched sample mean		$\widehat{\tau}_{\text{NATE}}$	$H_0 : \tau_{\text{NATE}} = 0$
	Treated	Control		One-sided p-value
Committing a crime	0.059	0.053	0.006	0.229

- How much bias from a hidden covariate would need to be present to mask an actual treatment effect?

Sensitivity analysis using a near-equivalence test

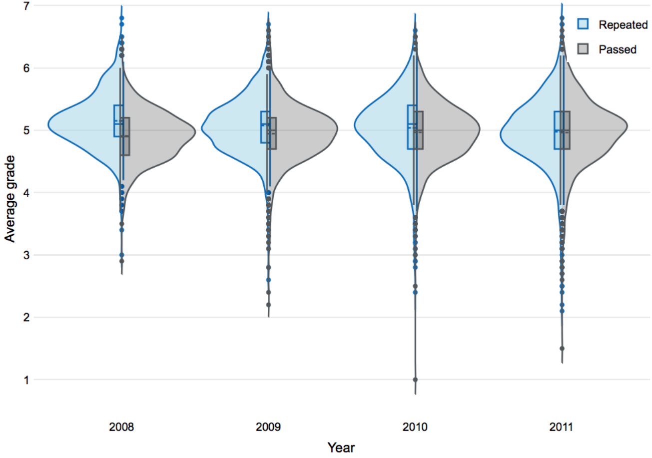
- In the absence of hidden bias, grade retention does not cause committing a juvenile crime

Table: NATE estimate

Outcome variable	Matched sample mean		$\widehat{\tau}_{\mathrm{NATE}}$	$H_0 : \tau_{\mathrm{NATE}} = 0$
	Treated	Control		One-sided p-value
Committing a crime	0.059	0.053	0.006	0.229

- How much bias from a hidden covariate would need to be present to mask an actual treatment effect?
- Two students matched for their covariates could differ in their odds of repeating by almost 33% before masking an effect of 4.3%

Displaying the outcomes



Seeing more from discontinuities

- If discontinuities are a keyhole into causality, can we see more?

Seeing more from discontinuities

- If discontinuities are a keyhole into causality, can we see more?
  - Heterogeneity?
  - Generalization?
  - Mediation?
  - Personalization?

Outline

Toward unit-level causal inferences

- How to evaluate the effects of a policy change on a city, region, country, or other aggregate unit?

Synthetic controls

- Proposed by Abadie and Gardeazabal (2003)
  - Economic cost of terrorist conflict in the Basque Country?
- A very creative, helpful, and transparent tool, often used to evaluate the effects of infrequent events, such as large-scale policy interventions, at an aggregate level of units
  - For example, countries, regions, cities
- As Abadie (2021) argues, even in experimental settings unit-level interventions may be infeasible (due to fairness constraints) or impractical (due to spillover effects)
  - In what follows, we will closely follow the work by Abadie and coauthors

Varied areas of application

- Immigration policy
- Tax programs
- Organized crime
- Tobacco control
- Localized lockdowns

Setting

- Single (typically, aggregate) treated unit and several untreated groups
- Data often has a strong time component, before and after the treatment
  - Pre-treatment data $t = 1, ..., T_0$
    - T0 : beginning of the treatment
  - Post-treatment data $t = T_1, ..., T$
- j indexes the units (e.g., country, state, city)
  - $j = 1$, treated
  - $j = 2, ..., J + 1$ untreated, so there are J potential controls or "donors"

Estimand

- Y_{jt}^{I}: potential outcome of unit j, exposed at $t = T_0 + 1, ..., T$
- Y_{jt}^{N}: potential outcome of unit j, not exposed
- We wish to estimate

$$\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$$

where $Y_{1t}^{I} = Y_{1t}$

Finding a good comparison

- Ideal: compare the exposed unit's outcome after the exposure to its outcome without the exposure
- Problem: counterfactuals; we never observe the exposed unit's outcome without the exposure
- Example: comparative case studies; compare, e.g., one country after the exposure to another country that is "comparable"
- Approach: finding a good comparison unit; there may be several other countries, but none alone serves as a valid contrast

Idea

- Can we combine potential comparison units to build a good comparison?
- The synthetic control method formalizes this idea by weighting potential comparison units so that, together and after weighting, they provide a good comparison

Weighted comparisons

- How might we determine whether a comparison unit is a good candidate?
- Often, we'd want it to look like the exposed unit before the exposure
- Seek weights such that the weighted comparison units "look like" the exposed unit before the exposure

How to choose the synthetic control weights?

- Roughly, the method finds weights that minimize a measure of pre-exposure covariate imbalance between the exposed and unexposed units
- For example, so that the weighted controls pre-exposure outcome trajectory is as close as possible to the exposed unit's pre-exposure outcome trajectory
- We can incorporate covariates into the weighting scheme as well (i.e., balance pre-exposure covariates), giving them different importance

Technical details

- To build the synthetic control, we find the weights $\boldsymbol{W} = (w_2, ..., w_{J+1})'$ that solve

$$\min ||\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}||$$

subject to

$$w_j \geq 0 \;\text{ for }\; j = 2, ..., J+1$$

and

$$w_2 + ... + w_{J+1} = 1$$

- where
  - $\boldsymbol{X}_1$ is the $k \times 1$ vector of pre-treatment covariates of the treated unit
  - $\boldsymbol{X}_0$ is the $k \times J$ matrix of pre-treatment covariates of the control units

Typical covariate distance

- Following Abadie (2021), typically

$$||\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}|| = \left\{ \sum_{h=1}^{k} v_h (X_{h1} - w_2 X_{h2} - \ldots - w_{hJ+1} X_{h2})^2 \right\}^{1/2}$$

where $v_1, \ldots, v_k$ capture the strength of the association between each of the k covariates and Y_{1t}^N

Synthetic control estimator

- We use the weights in the linear estimator

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

A bias bound

- Abadie et al. (2010) provide a bias bound under the factor model

$$Y_{1t}^N = \boldsymbol{\theta}_t \boldsymbol{X}_i + \boldsymbol{\lambda}_t \boldsymbol{U}_i + \varepsilon_{it}$$

where $\boldsymbol{X}_i$ are observed covariates, $\boldsymbol{U}_i$ are unobserved covariates, and ε_{it} is a random error term

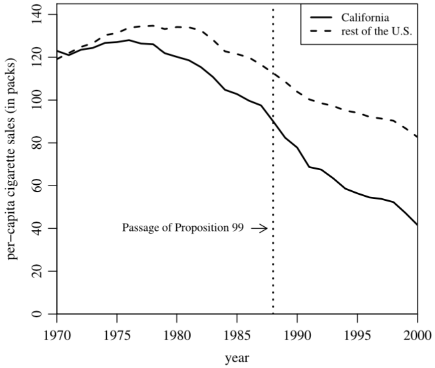
Impact of California's tobacco control program

- Does tobacco control reduce smoking?
- Proposition 99
  - Implemented in California in 1989
  - Increased cigarette excise tax by a large amount
  - Required funds to be used for anti-smoking education and clean indoor air ordinances

Data

- Annual measurements for California and 36 untreated states from 1970 to 2000
- Outcome: annual cigarette packs sold per capita
- Covariates include per capita beer consumption and log income, retail price of a cigarette pack, and percentage of population aged 15-24 years

Unadjusted trends in per-capita cigarette sales (Abadie et al., 2010)



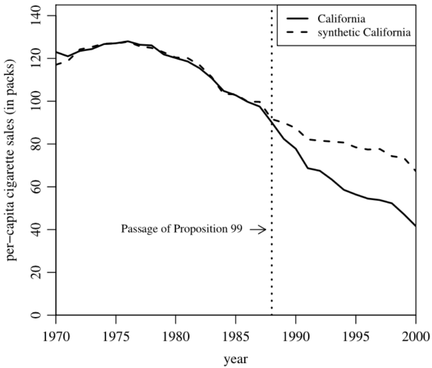
Diagnostics

- Diagnostics are important to evaluate the implementation of the synthetic control method
- Assess covariate balance:
  - How well the outcome trajectory of the exposed unit is approximated by the weighted unexposed units?
  - How well are other pre-exposure covariates balanced?
- Look at the weights:
  - How sparse are the weights, so that only a few units actually contribute to the synthetic control?
  - Do the units that comprise the synthetic control make sense (e.g., are they geographically nearby)?

Cigarette sales predictor means (Abadie et al., 2010)

Variables	California Real	California Synthetic	Average of 38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Adjusted trends in per-capita cigarette sales (Abadie et al., 2010)



State weights in the synthetic California (Abadie et al., 2010)

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

Comparing weights

- How do the synthetic control weights compare to the following?
  - Matching weights
  - Minimal weights
  - Regression weights
    - Linear “machine learning”
- Relevant dimensions:
  - Sparsity
  - Dispersion
  - Extrapolation
  - Interpretation

Geometric interpretation of the synthetic weights

- The synthetic control: a linear combination of the potential controls
- If $\boldsymbol{X}_1$ does not belong to the convex hull of $\boldsymbol{X}_0$, the synthetic control is unique and sparse
- If $\boldsymbol{X}_1$ belongs to the convex hull of $\boldsymbol{X}_0$, the synthetic control may not be unique nor sparse

A penalized synthetic control approach (1)

- Abadie and L'Hour (2020)

$$\min ||\boldsymbol{X}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{X}_j|| + \lambda \sum_{j=2}^{J+1} w_j ||\boldsymbol{X}_1 - \boldsymbol{X}_j||^2$$

subject to

$$w_j \geq 0 \;\text{ for }\; j = 2, ..., J+1$$

and

$$w_2 + ... + w_{J+1} = 1$$

A penalized synthetic control approach (1)

- Abadie and L'Hour (2020)

$$\min ||\boldsymbol{X}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{X}_j|| + \lambda \sum_{j=2}^{J+1} w_j ||\boldsymbol{X}_1 - \boldsymbol{X}_j||^2$$

subject to

$$w_j \geq 0 \text{ for } j = 2, ..., J+1$$

and

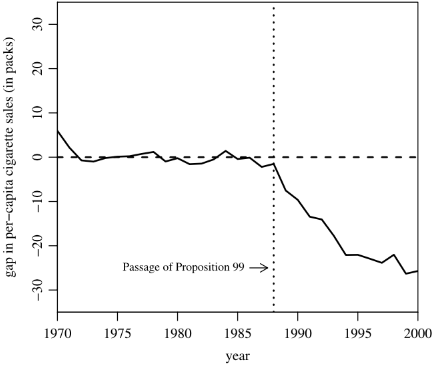
$$w_2 + ... + w_{J+1} = 1$$

- Extreme cases:
  - $\lambda = 0$: pure synthetic control
  - $\lambda = \infty$: nearest neighbor matching
- $\lambda > 0$ reduces the extent of interpolation bias

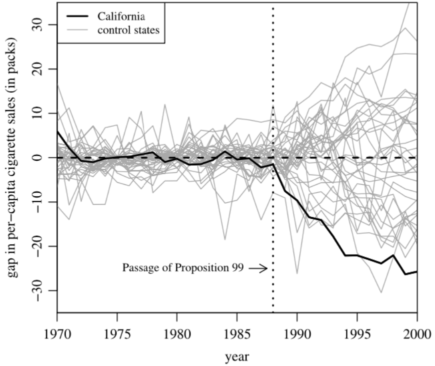
Inference and stability

- Permutation inference: permute which unit is labeled as the exposed unit and get the distribution of a test statistic
- Placebo test: backdate the intervention and test whether we still get an effect
- Robustness check: leave-one-out of the control pool and see if estimates change much

Sales gap between California and synthetic California (Abadie et al., 2010)



Synthetic California and synthetic placebo gaps in other control states (Abadie et al., 2010)



Data requirements

- Sufficient covariates for a given outcome and intervention
  - Necessary for no unmeasured confounding cross-sectionally (build a credible synthetic comparison)
- Sufficient pre-intervention information
  - Necessary for no unmeasured confounding longitudinally (steadily track the trajectory of the outcome before the intervention)
- Sufficient post-intervention information
  - Necessary to understand the timing of the effects (assess gradual/abrupt, sustained/dissipative effects)

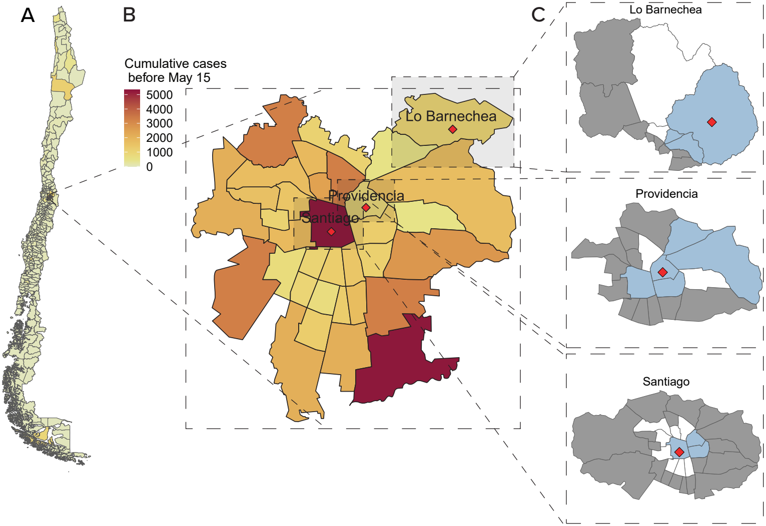
Another case study: impact of localized lockdowns in the COVID-19 pandemic

- Essential to control viral transmission early in the pandemic
  - Examples: face masks, physical distance
- Localized lockdowns in small geographic areas
  - Lower social and economic costs than interventions in larger areas
- At the time, limited evidence on their effectiveness
  - Impact of duration and spillover effects

Causal effects of localized lockdowns

- Direct effects of lockdowns in a municipality
- Spillover effects of lockdowns in neighboring municipalities
- Total effects: sum of direct and spillover effects
    - Effects further modulated by the duration of the lockdown
- We estimated these effects in Chile

Chile, Greater Santiago, three municipalities

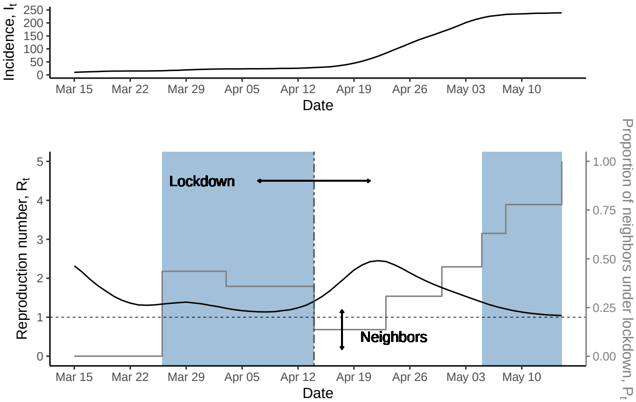


Integrated data from three sources

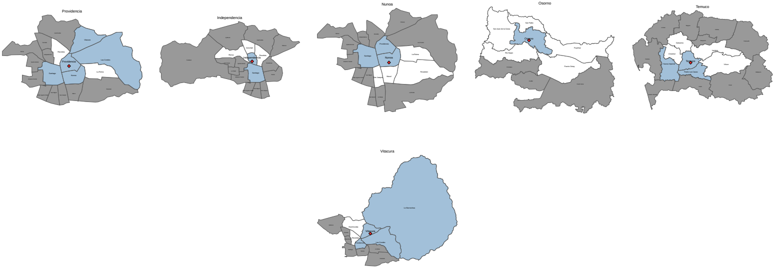
- Epidemiologic records: Ministry of Health
  - COVID-19 cases defined as PCR-confirmed SARS-CoV-2 infections
- Household data: National Socioeconomic Characterization Survey
- Population measurements: National Census

Study question

- Impact of localized lockdowns?
  - Causal effect of the duration and the proportion of the neighboring population under lockdown on the incidence and reproduction number



Building synthetic or “clone” municipalities



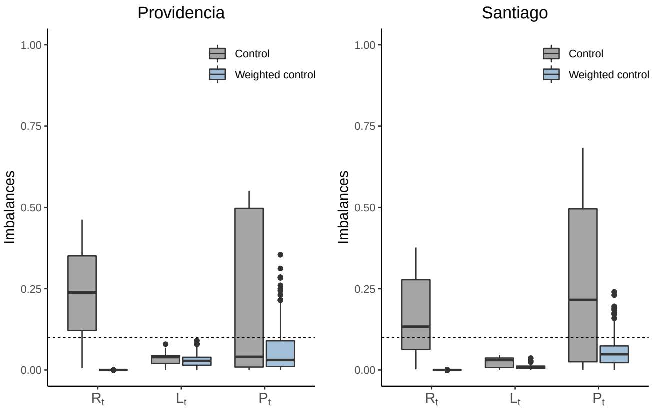
Variables

- Covariates:
  - X_i: pre-intervention baseline covariates in municipality i
  - $L_{i[t_h-h,t_h-1]}$, $P_{(i)[t_h-h,t_h-1]}$, and $R_{i[t_h-h,t_h-1]}$: "histories" of...
- Interventions:
  - L_{it}: proportion of population under lockdown in municipality i
  - $P_{(i)t}$: proportion of population under lockdown in the neighbors of municipality i
- Outcomes:
  - I_{it}: incidence in municipality i at time t
  - R_{it}: instantaneous reproduction number of municipality i

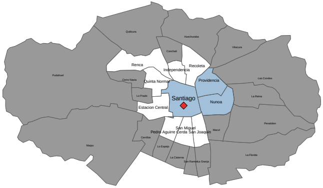
Covariate balance: Santiago

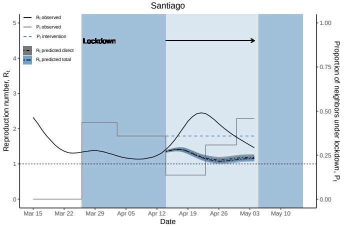
Baseline covariates	Santiago	Control Set	Synthetic Santiago
Rural	0.00	0.03	0.00
Female	0.50	0.52	0.50
Over 65	0.07	0.11	0.07
Poverty	0.04	0.07	0.04
Overcrowding	0.19	0.12	0.19
Poor sanitation	0.02	0.04	0.02
Income	593633.86	375521.94	593633.86
Area (small)	1.00	0.36	1.00
Area (medium)	0.00	0.35	0.00
Area (large)	0.00	0.30	0.00
Lagged variables	Santiago	Control Set	Synthetic Santiago
R_{t-7}	1.14	1.13	1.14
R_{t-6}	1.14	1.12	1.14
R_{t-5}	1.14	1.10	1.14
R_{t-4}	1.17	1.09	1.17
R_{t-3}	1.19	1.08	1.19
R_{t-2}	1.24	1.07	1.24
R_{t-1}	1.31	1.07	1.31
L_{t-7}	1.00	0.95	1.01
L_{t-6}	1.00	0.95	1.01
L_{t-5}	1.00	0.96	1.00
L_{t-4}	1.00	0.96	1.00
L_{t-3}	1.00	0.96	1.00
L_{t-2}	1.00	0.96	0.99
L_{t-1}	1.00	0.96	0.99
P_{t-7}	0.36	0.74	0.35
P_{t-6}	0.36	0.75	0.33
P_{t-5}	0.36	0.76	0.34
P_{t-4}	0.36	0.76	0.36
P_{t-3}	0.36	0.77	0.38
P_{t-2}	0.36	0.78	0.36
P_{t-1}	0.36	0.79	0.39

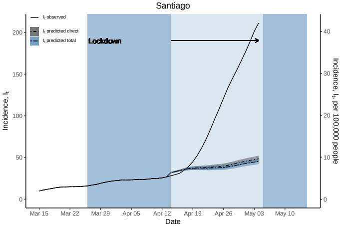
Distribution of imbalances



Direct and total effects: Santiago



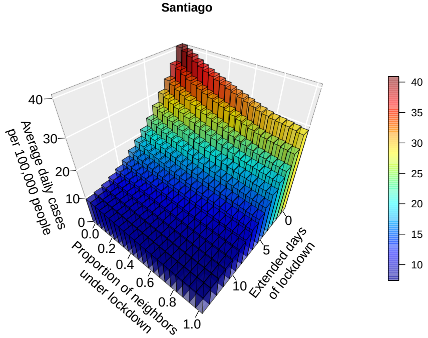


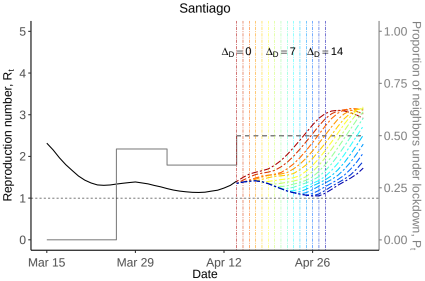


Results

- These effects represent 33-62% reductions in reported cases in that time frame
- The reductions would have been even larger if it was possible to control lockdowns in neighbouring municipalities

Duration and spillovers, $\Delta_D = 0, ..., 14$, $P_t \in [0, 1]$





Takeaways

- The effects of localized lockdowns are strongly modulated by their duration and are affected by indirect effects from neighboring geographic areas that are not under lockdown
- Extending localized lockdowns slow down the epidemic; however, by themselves, localized lockdowns are insufficient to control epidemic growth due to indirect effects from neighboring areas

Big picture: strengths of the synthetic control method

- Transparency of fit: we can clearly evaluate how well the synthetic control mimics the exposed unit's pre-exposure outcome evolution
- No extrapolation: most implementations require that the weights are positive, which restricts the estimate to be an interpolation
- Separation of design and analysis stages: the weights are obtained using pre-treatment measurements, so there is less risk of fishing
- Flexibility: we can use the same weights to evaluate effects on different timescales (short-, long-run)
- Simplicity in estimation: we can compute a simple weighted average of the outcomes

Questions

- Other ways to make inferences?
- Sensitivity analyses?
- How to study mediation?
- Manage interpolation and extrapolation bias?

Causal Inference
Designs

José R. Zubizarreta
Harvard University

09/04/2023
CUSO Doctoral School in Statistics and Applied Probability
Saignelégier, Switzerland