

Difference-in-Differences

Impact evaluation: Effects of social policies on health

22 Feb 2022

Sam Harper & Arijit Nandi



McGill

Department of
Epidemiology, Biostatistics
and Occupational Health

CONTENTS

I. OVERVIEW

II. ESTIMATING DD EFFECTS

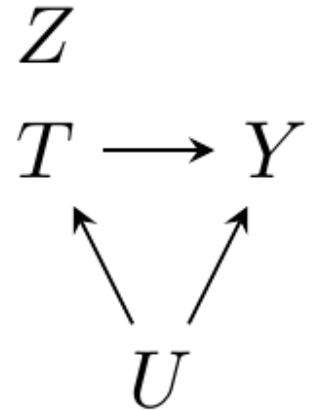
III. EXTENSIONS

IV. ROBUSTNESS CHECKS

V. STAGGERED TREATMENTS

I. OVERVIEW

Thinking about research design

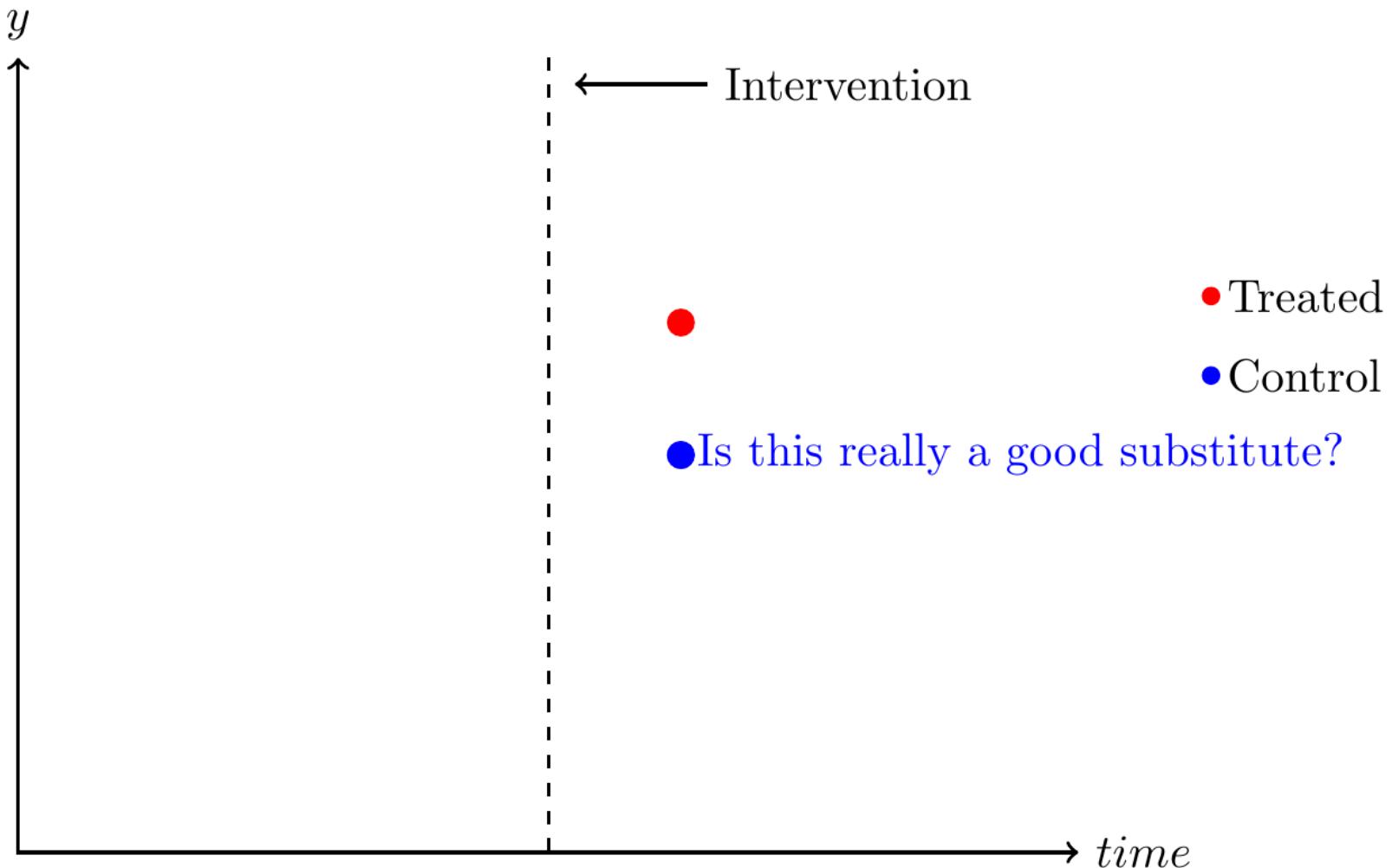


- Without randomization (Z), Approaches using quasi-experiments focus on exploiting:
 1. A treatment group that experiences a **change** in the exposure of interest.
 2. Comparison with an appropriate control group that does not experience a change in exposure.
- Recall the potential outcomes framework. We need a substitute population (treated and controls):

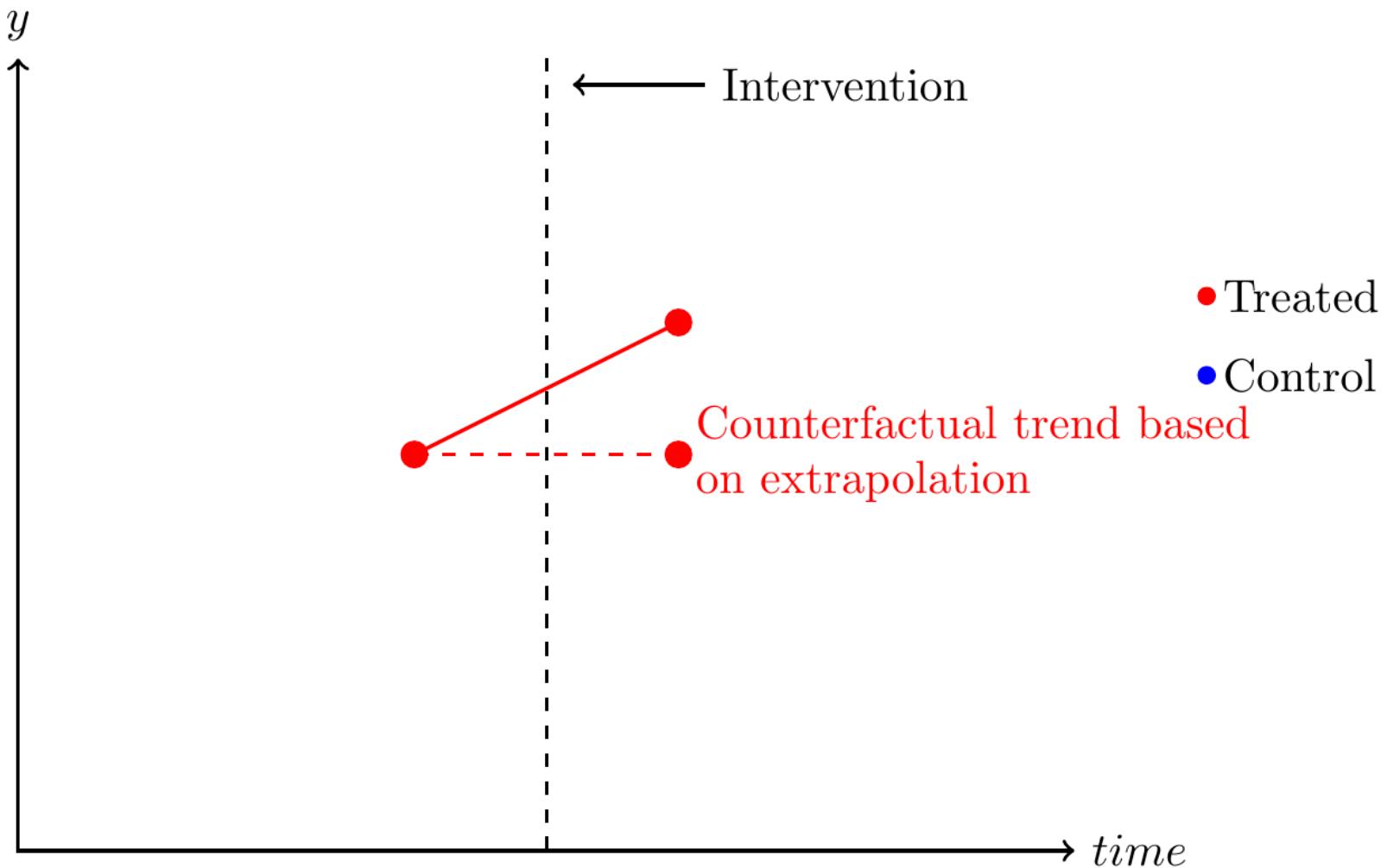
$$E[Y^1 - Y^0] = E[Y^1|T = 1] - E[Y^0|T = 0]$$

- Where should we get our counterfactual?

One-group posttest design with control group



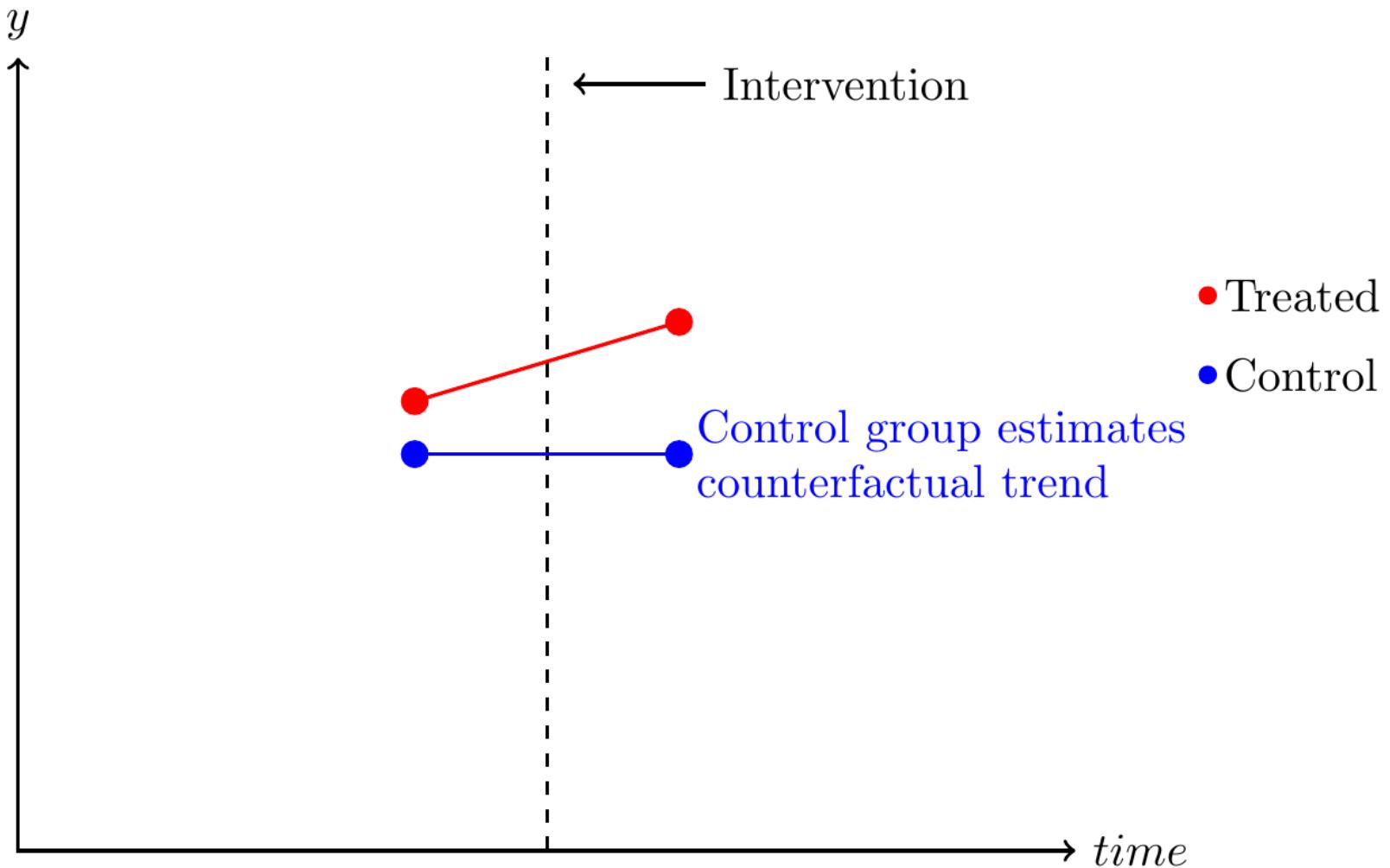
One-group pretest-posttest design



One-group pretest-posttest design

- Even a single pretest observation provides some improvement over the posttest only design.
- Now we derive a counterfactual prediction from the same group before the intervention.
- Provides weak counterfactual evidence about what would have happened in the absence of the program.
 - We know that Y_{t-1} occurs before Y_t (correct temporal ordering).
 - Could be many other reasons apart from the intervention that $Y_t \neq Y_{t-1}$.
- Stronger evidence if the outcomes can be reliably predicted and the pre-post interval is short.
- Better still to add a pretest and posttest from a control group.

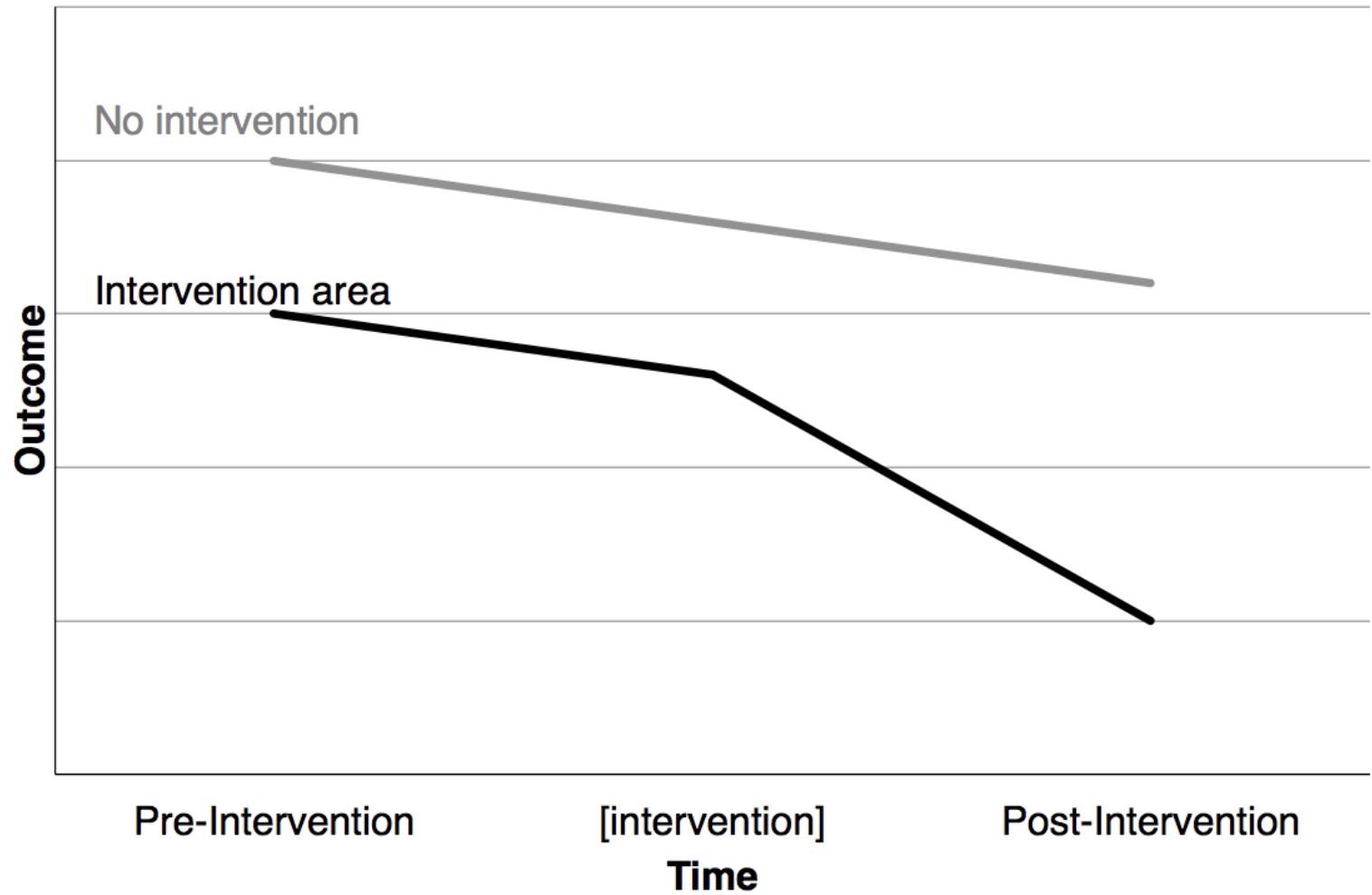
Adding pretests for both groups



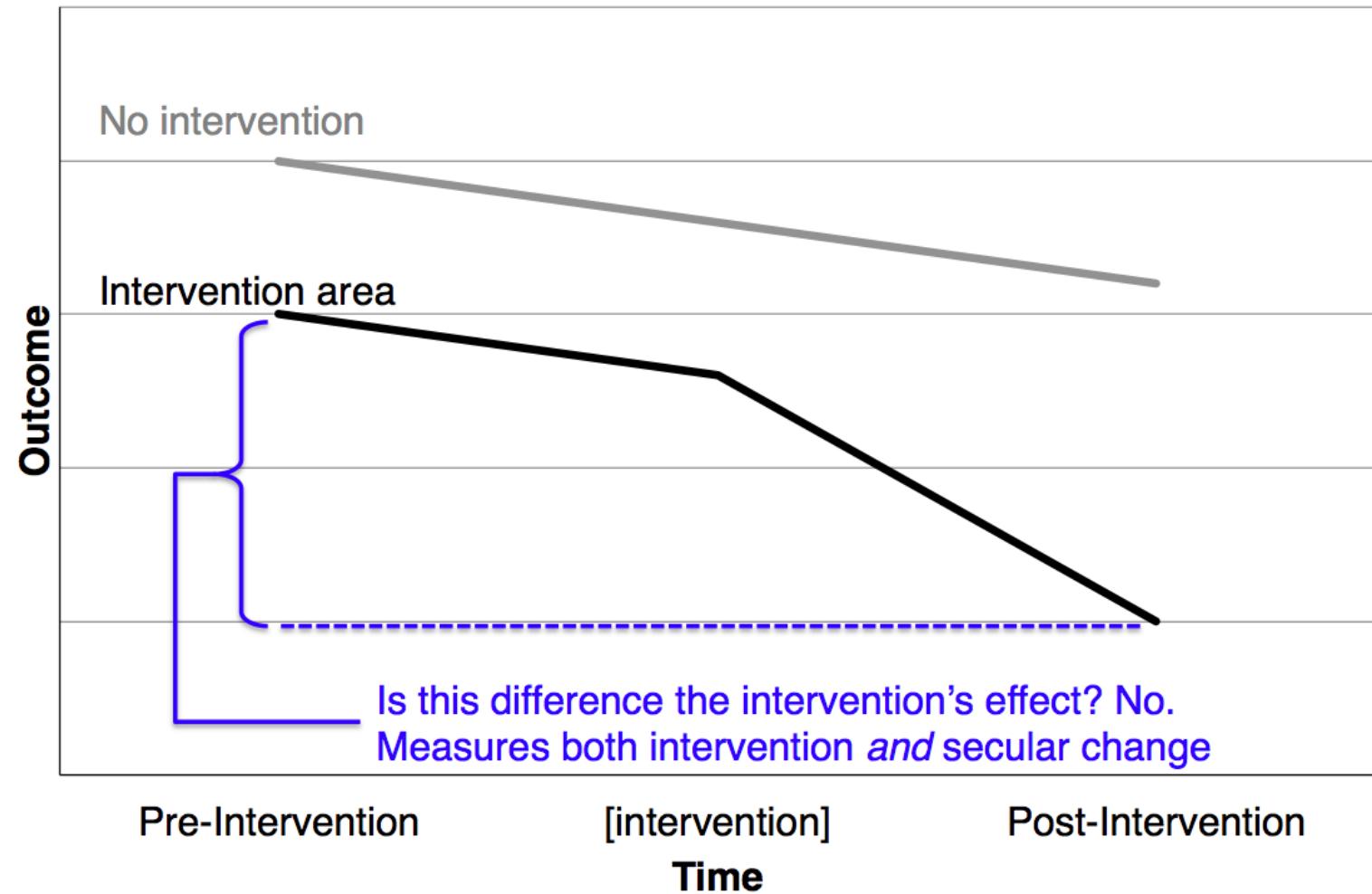
How does this help?

- Pre/post in control helps resolve this by differencing out any **time-invariant** characteristics of both groups.
 - Many observed factors don't change over the course of an intervention (e.g., geography, parents' social class, birth cohort).
 - Any time-invariant *unobserved* factors also won't change over intervention period.
 - We can therefore effectively control for them.
- Measuring same units before and after a program cancels out any effect of all of the characteristics that are unique to that observation and that do not change over time.
- This also has the benefit of canceling out (or controlling for) unobserved time-invariant characteristics.

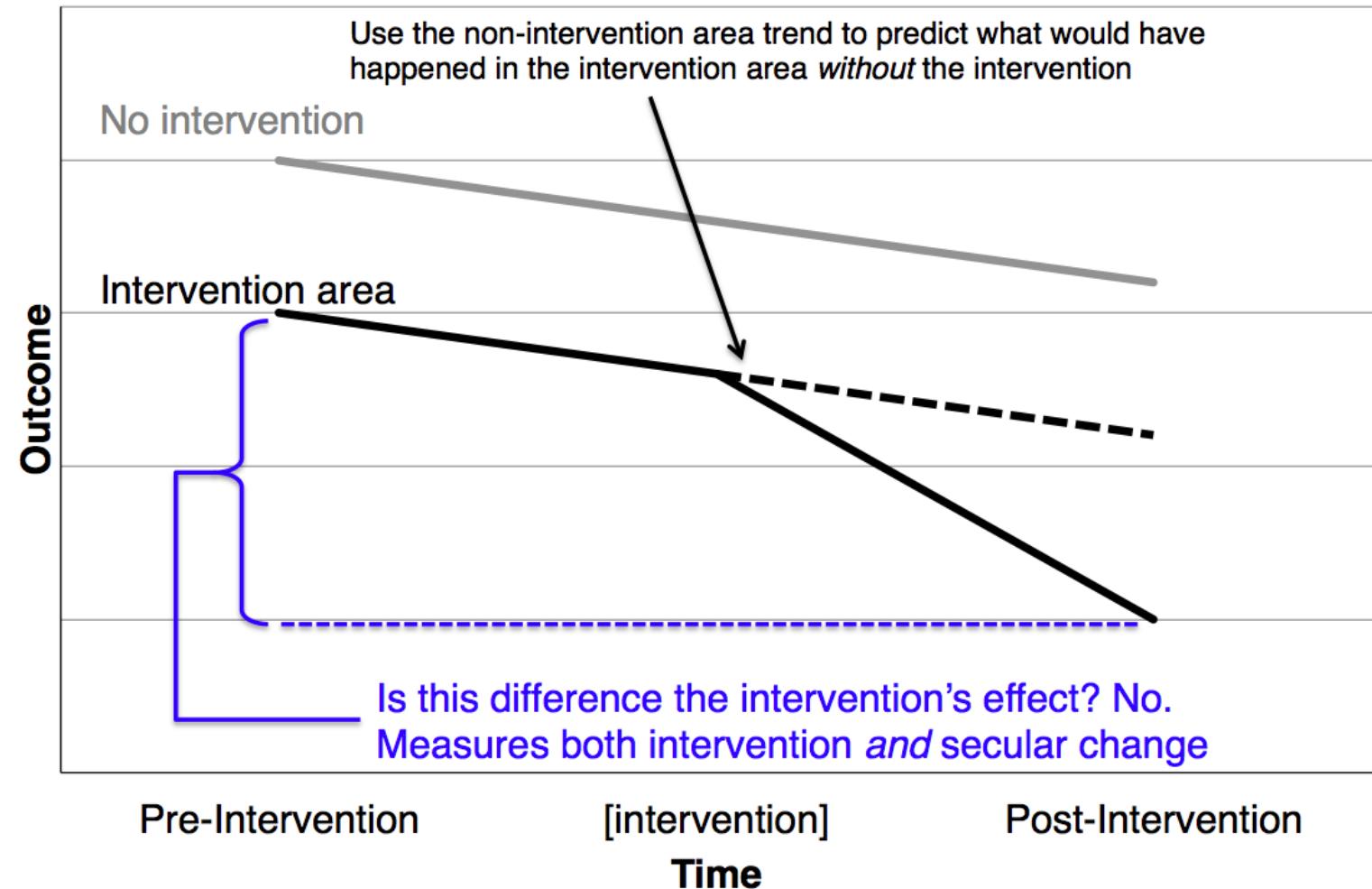
The need for a control group



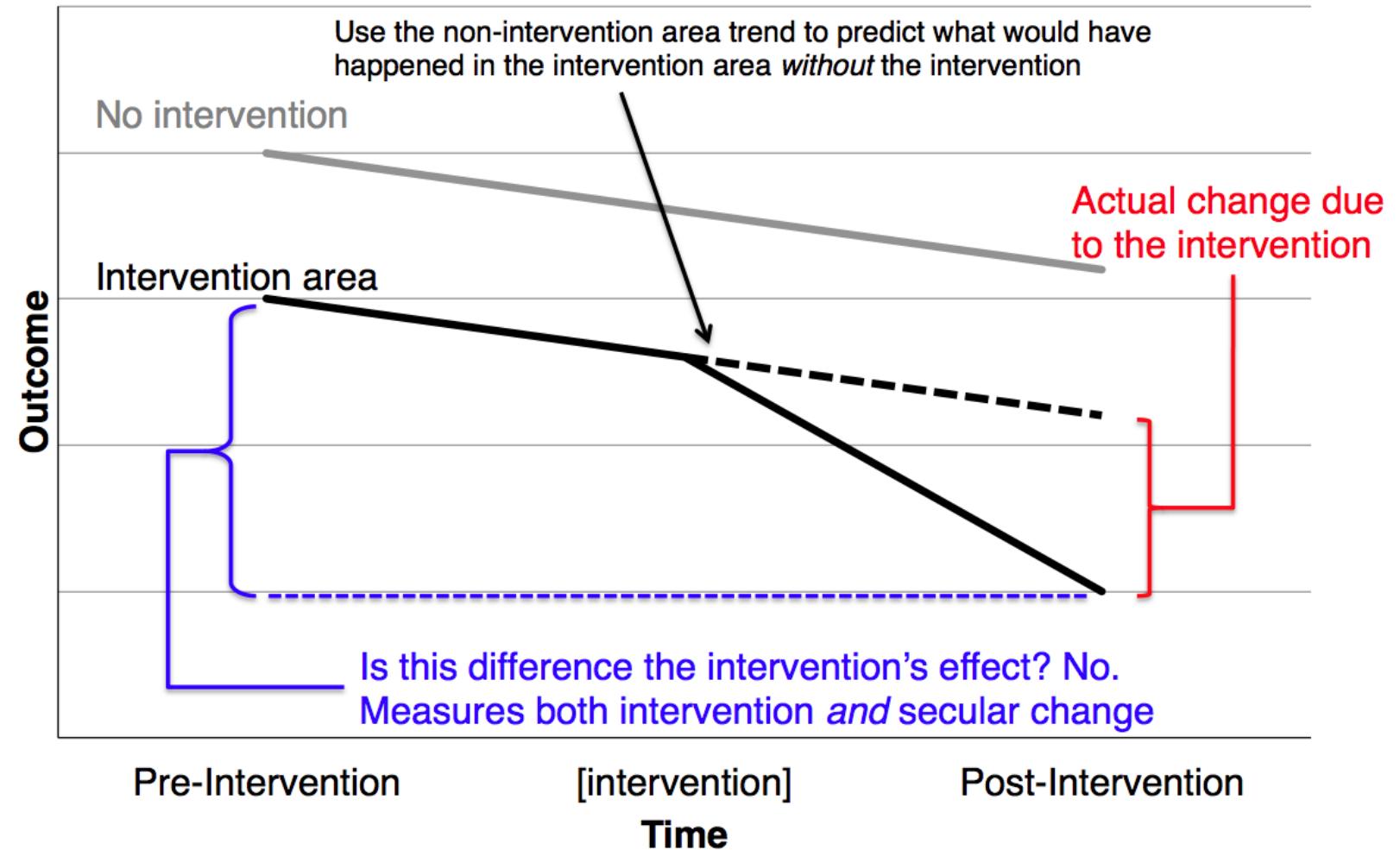
What would
have
happened
without the
intervention?



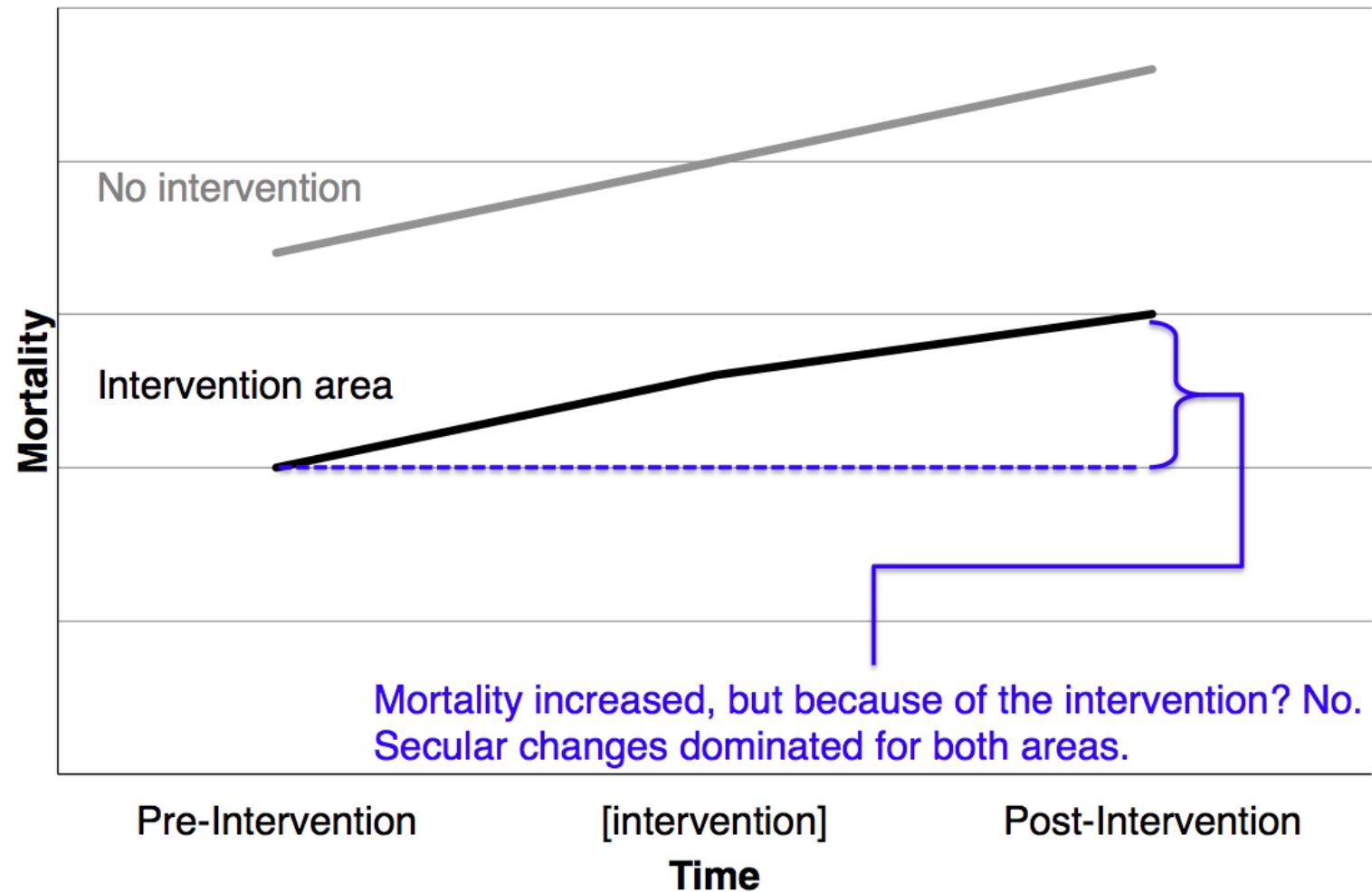
What would have happened without the intervention?



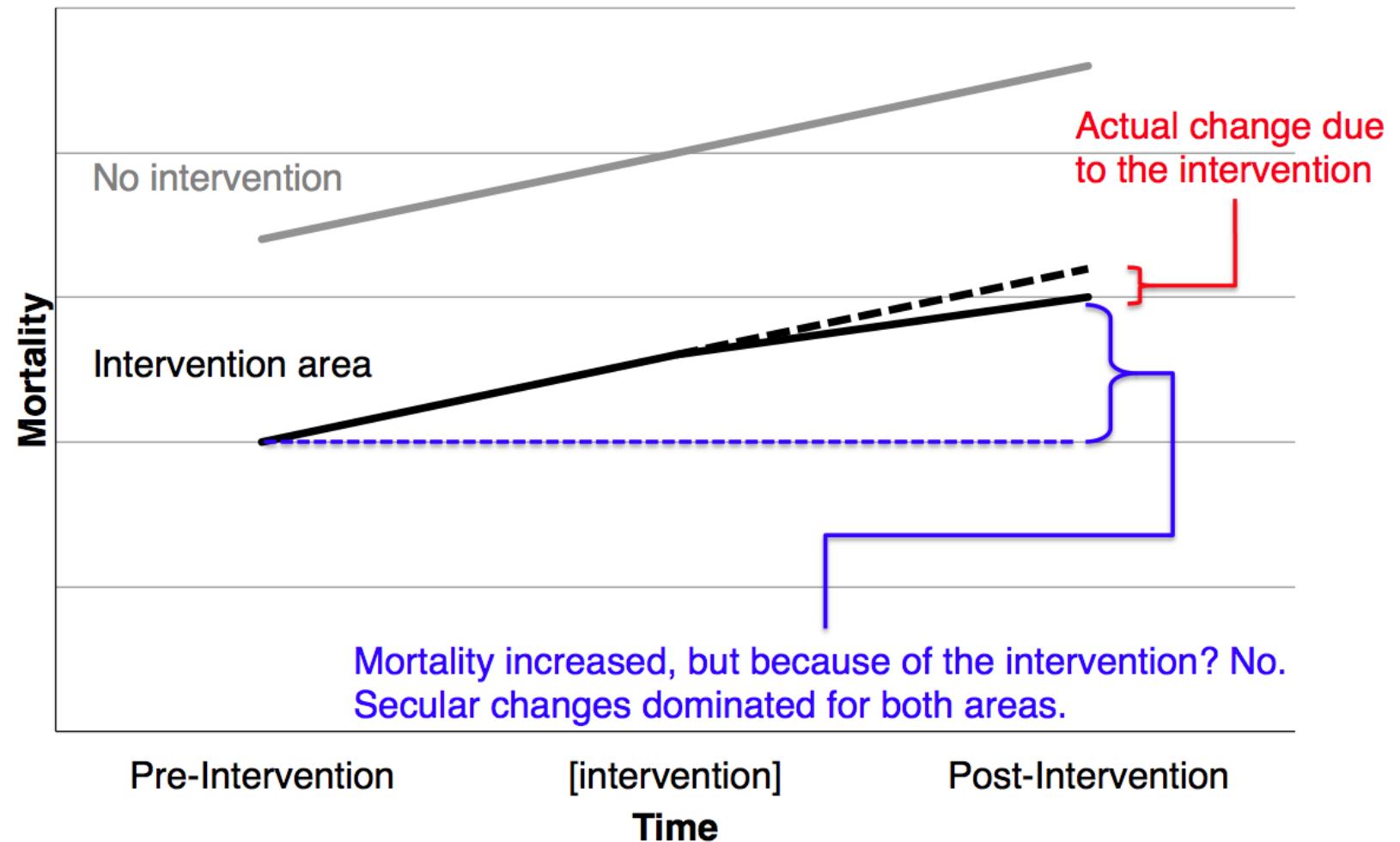
What would have happened without the intervention?



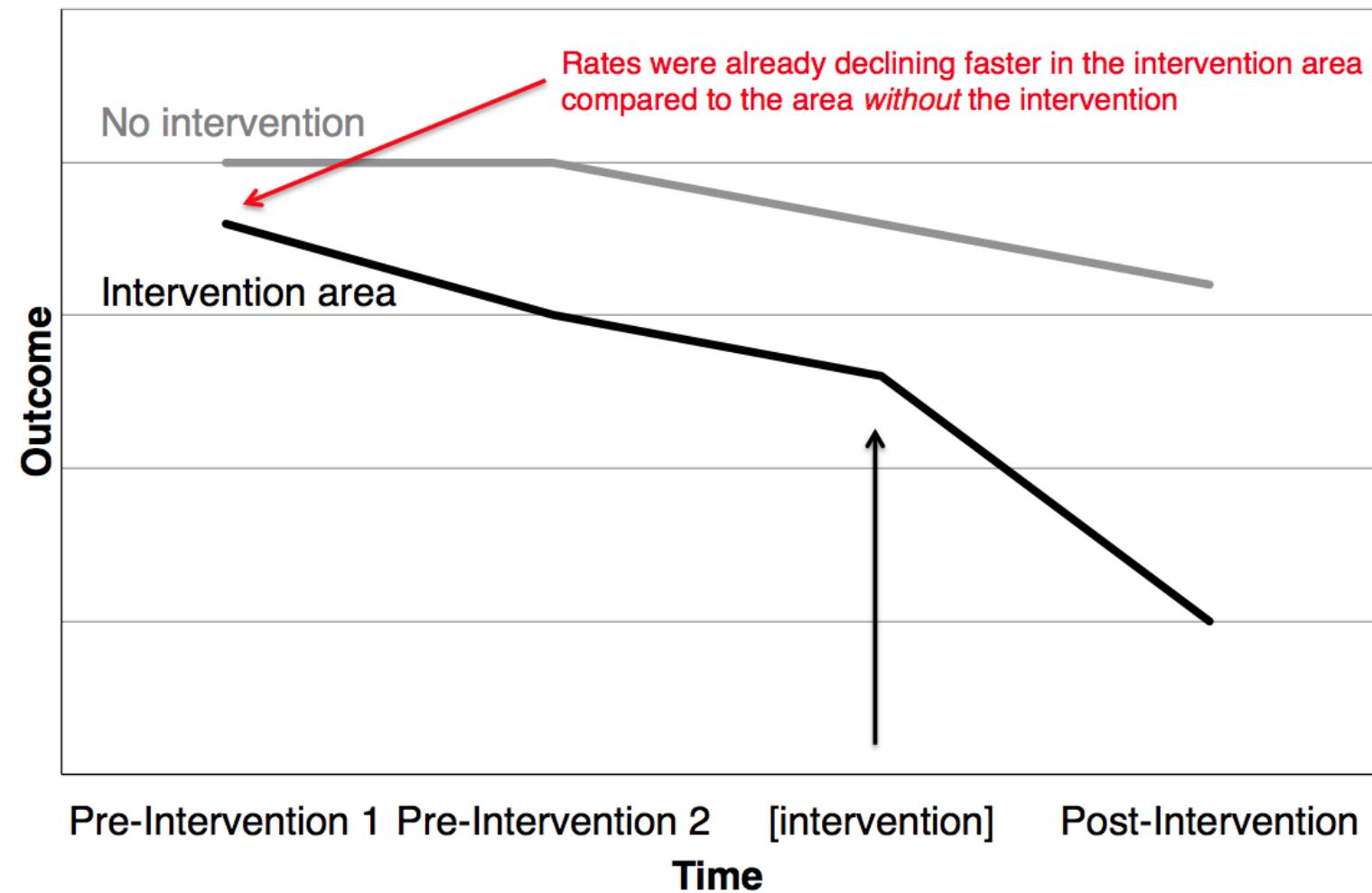
What if things worsened over time?



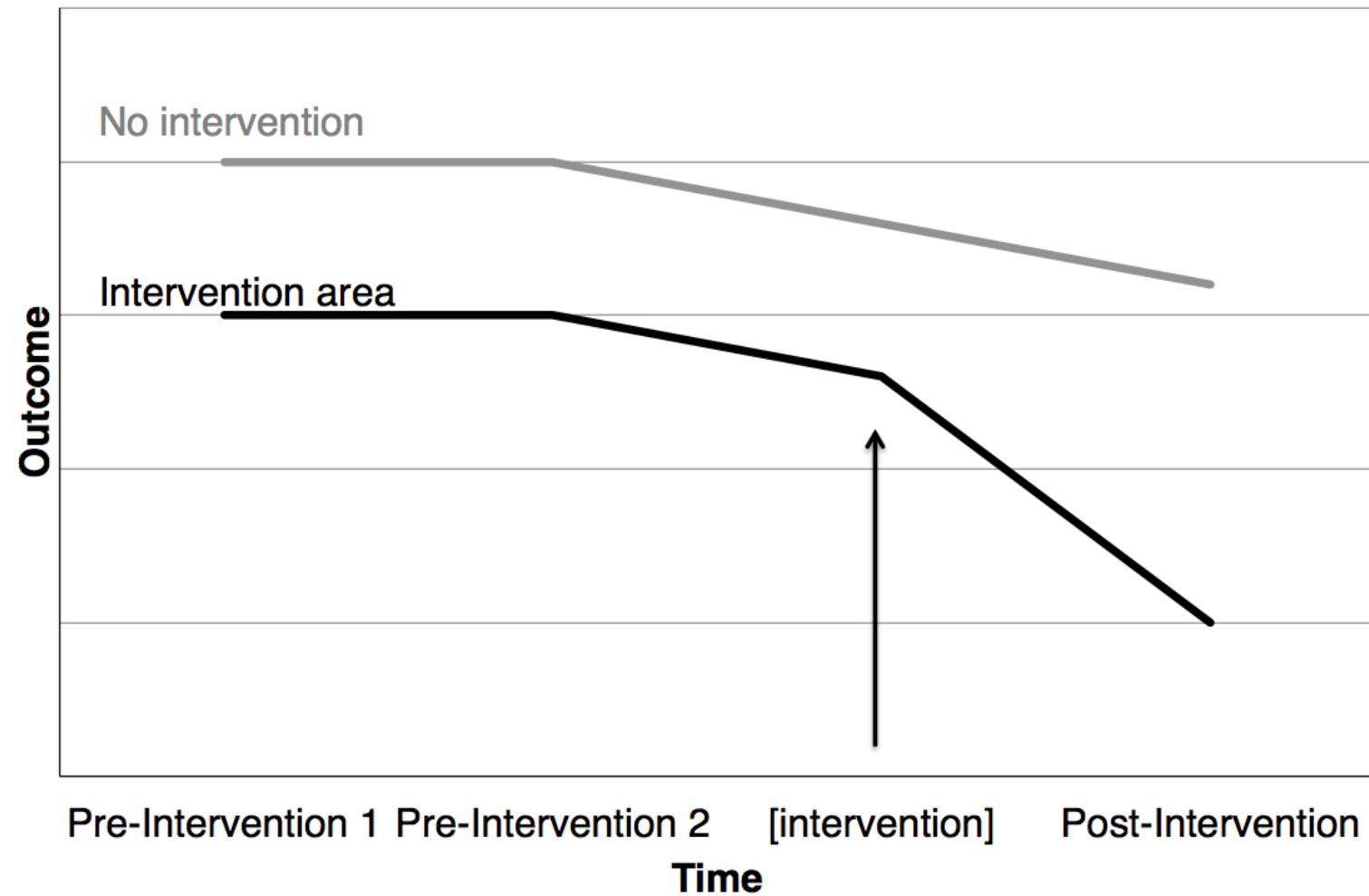
What if things worsened over time?



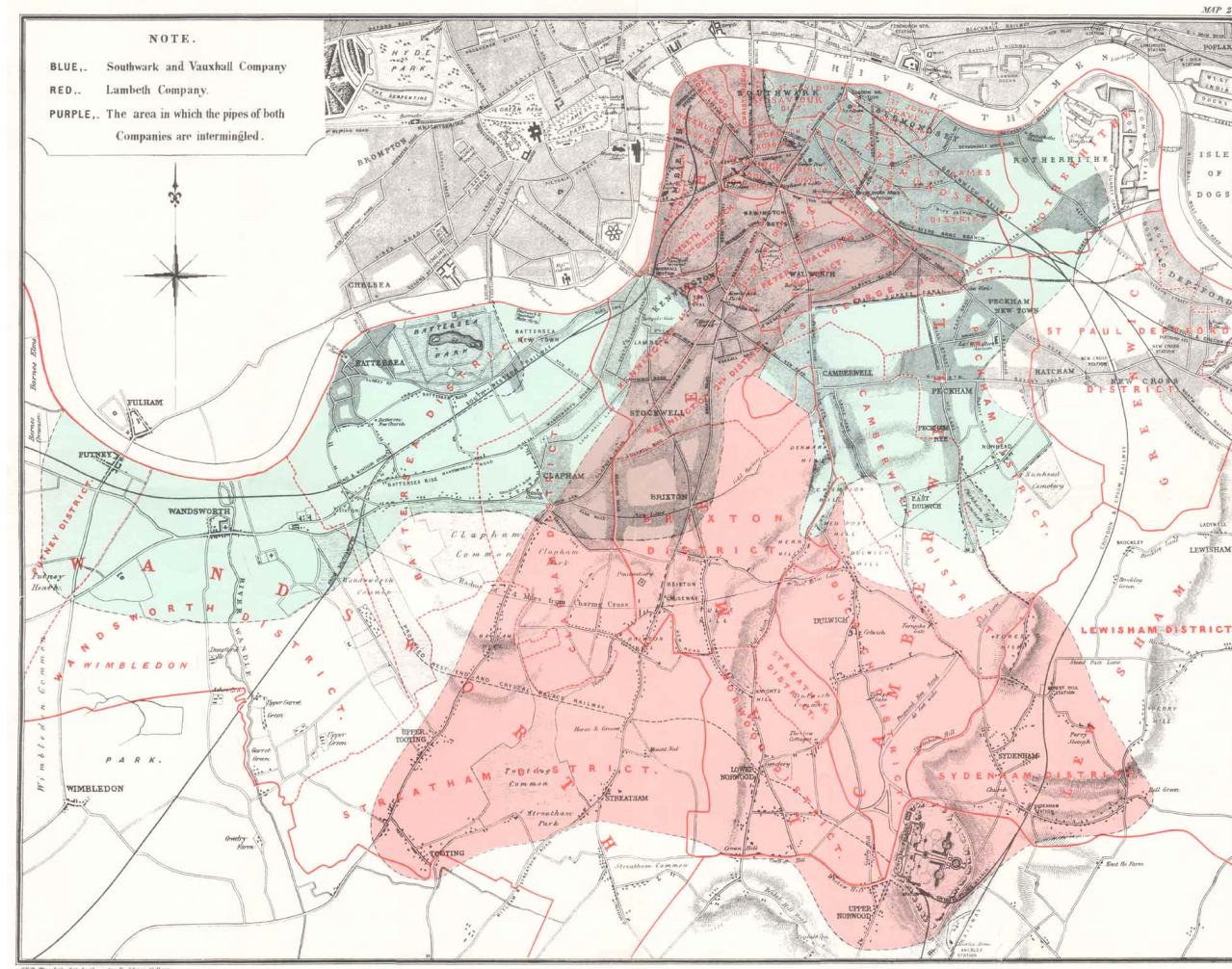
More time periods are better for evaluation



Parallel pre-intervention trends increase "exchangeability"



Classic example from epidemiology: Water and cholera



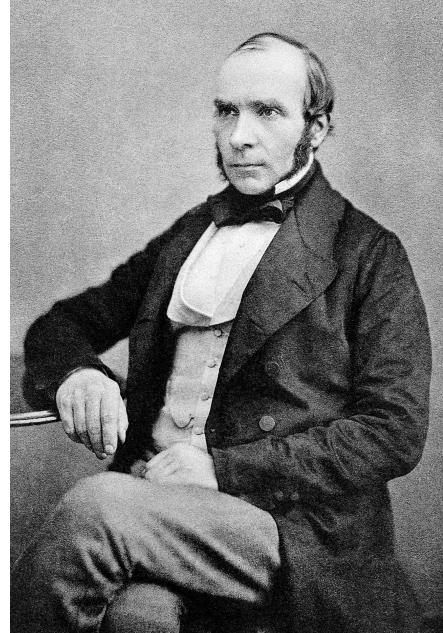
See Snow (1855) reprinted as Snow, Frost, and Richardson (1965)

Snow's method

- Couldn't randomize.
- Knew Lambeth required to move intake upstream of London *after* 1849.
- SV did not move, and communities were similar in many ways.
- Used SV as an 'unaffected' control community.
- Did not estimate DD parameter, but idea was there.

Region	Rate (1849)	Rate (1854)	Post-Pre
Lambeth (treated)	130.1	84.9	-45.2
Southwark + Vauxhall (control)	134.9	146.6	11.7
Group Diff (treat - control)	-4.8	-61.7	-56.9

Why is Snow's work compelling?



- Evidence of pre-treatment equivalence between groups:

"In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different companies..."
- Treatment groups lacked knowledge of mechanisms, or intervention:

"divided into two groups without their choice, and, in most cases, without their knowledge"

See Snow (1855) reprinted as [Snow, et al. \(1965\)](#) and also [Freedman \(1991\)](#).

Other recent examples

Effect of US Deferred Action for Childhood Arrivals (DACA) immigration program on health outcomes.

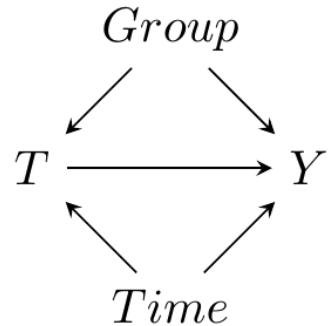
- Compared changes in health outcomes among individuals who met key DACA eligibility criteria (based on age at immigration and at the time of policy implementation) *before and after* program implementation versus changes in outcomes for individuals who did not meet these criteria.

Effect of changing the legal age of handgun purchases and adolescent suicide in US

- Compared changes in suicide rates in US states that *changed* the age at which individuals could legally purchase handguns (both increases and decreases) to US states that did not change the age at which handguns could be purchased.

See [Venkataramani, Shah, O'Brien, Kawachi, and Tsai \(2017\)](#) 30047-6 for the DACA study and [Raifman, Larson, Barry, Siegel, Ulrich, Knopov, and Galea \(2020\)](#) for the suicide study.

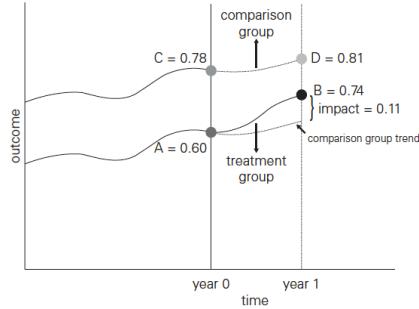
Difference-in-Differences: Basic Idea



The simplest DD setting:

- Outcomes are observed for units observed in one of two groups:
 - Treated vs. Control
- Outcomes observed in one of two time periods.
 - Before and After the intervention.
- Treated: only units in one of the two groups are exposed to a treatment, in the second time period.
- Control: Never observed to be exposed to the treatment.

Difference-in-Differences: Basic Idea



- The average change over time in the non-exposed (control) group is subtracted from the change over time in the exposed (treatment) group.
- Double differencing removes biases in second period comparisons between the treatment and control group that could result from:
 - Fixed (i.e., non time-varying) differences between those groups.
 - Comparisons over time in the treatment group that could be the result of time trends unrelated to the treatment.

Key Assumption: Parallel Trends

- Basic DD controls for any time invariant characteristics of both treated and control groups.
- Does not control for any **time-varying** characteristics.
- If another policy/intervention occurs in the treated (or control) group at the same time as the intervention, we cannot cleanly identify the effect of the program.
- DD main assumption: in the absence of the intervention treated and control groups would have displayed similar **trends**.
- This is called the *parallel trends* assumption.

Impossible to verify (see [Gertler, Martinez, Premand, Rawlings, and Vermeersch \(2011\)](#))

II. ESTIMATING DD EFFECTS

Difference-in-Differences without Regression

- DD is just differences in means!

Let $\mu_{it} = E(Y_{it})$

- $i = 0$ is control group, $i = 1$ is treatment.
- $t = 0$ is pre-period, $t = 1$ is post-period.
- One 'difference' estimate of causal effect is: $\mu_{11} - \mu_{10}$ (pre-post in treated)

- Differences-in-Differences estimate of causal effect is:
$$(\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$$

Area	Before	After	Difference
Treated	135	100	-35
Control	80	60	-20
T - C	55	40	-15

DD Regression: Two Groups, Two Periods (2x2)

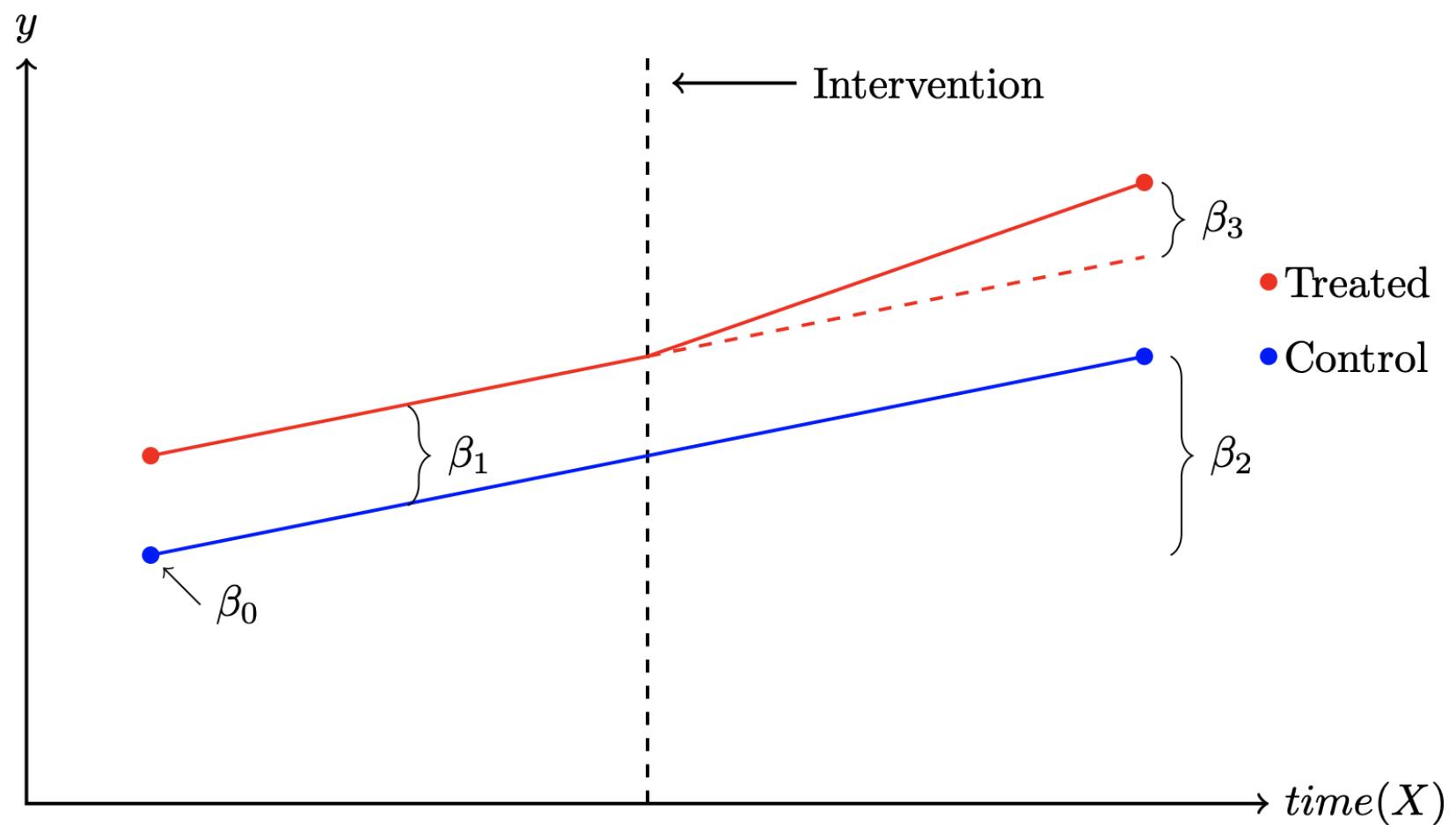
- Single treated and control group, two periods
- β_1 = Treated group
- β_2 = Post period
- β_3 = Product term

y	group	time	treat?	post?	treatXpost
:	1	1	0	0	0
:	1	2	0	1	0
:	2	1	1	0	0
:	2	2	1	1	1

$$Y = \beta_0 + \beta_1 * treat + \beta_2 * post + \beta_3 * treat * post$$

Visual interpretation of parameters from linear DD model

$$Y = \beta_0 + \beta_1 Treat + \beta_2 Post + \beta_3 Treat * Post + \varepsilon_t$$



Difference-in-differences (usually) estimates the ATT

- Our DD model is: $Y = \beta_0 + \beta_1 Treat + \beta_2 Post + \beta_3 Treat * Post + \varepsilon$.
 - we showed that β_3 is the DD estimate in the linear model case.
- In the (possibly counterfactual) absence of intervention, the expected outcome is:
 - $E(Y_i^0 | T = 1, A = 0) = \beta_1 + \beta_2$
- In the (possibly counterfactual) presence of intervention, the expected outcome is:
 - $E(Y_i^1 | T = 1, A = 1) = \beta_1 + \beta_2 + \beta_3$
- ATT is the expected difference in $Y_i^1 - Y_i^0$ for those treated in the post-period:
 - $ATT = E(Y^1 - Y^0 | T = 1) = \beta_3$

Simple Two Period, Two Group Example

Suppose Quebec passes an anti-poverty policy in 2019.

How to estimate
the impact?

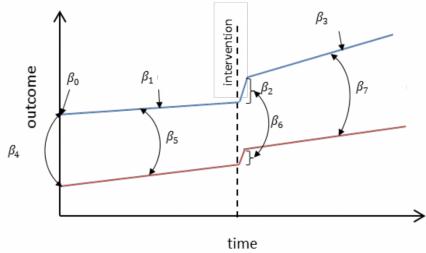
- $Treat = 1$ if Quebec, 0 if Ontario.
- $Post = 1$ if 2019, 0 if 2018.
- $Treat * Post = 1$ if Quebec in 2019, 0 otherwise.

$$Y = \beta_0 + \beta_1 Treat + \beta_2 Post + \beta_3 Treat * Post + \varepsilon$$

Province, Time	Estimate	Time Diff	DD
Ontario, 2018	β_0		
		$\} \beta_2$	
Ontario, 2019	$\beta_0 + \beta_2$		$\} \beta_3$
Quebec, 2018	$\beta_0 + \beta_1$		$\} \beta_2 + \beta_3$
Quebec, 2019	$\beta_0 + \beta_1 + \beta_2 + \beta_3$		

Wait ... isn't this just "Controlled" ITS (CITS)?

Remember this?



Well...kind of?

- Both designs use untreated groups to estimate the counterfactual for the treated group.
- Simple 2x2 DD study does not require any time series data.
- CITS often assesses impact on step change and slope.
- DD often estimates average effect in post period.
- Both can incorporate longer time series and dynamic effects.

See the exchange between [Benmarhnia and Rudolph \(2019\)](#) and [Lopez Bernal, Cummins, and Gasparrini \(2019\)](#) for more.

Reformulation of the model using 'fixed effects'

Express our earlier model using 'fixed effects':

$$Y = \beta_0 + \beta_1 * Group2 + \beta_2 * Time2 + \beta_3 * policy$$

- Dummy for Group
- Dummy for Time
- *Time-varying* policy indicator

y	group	time	treat?	post?	treatXpost	Group 2	Time 2	policy
:	1	1	0	0	0	0	0	0
:	1	2	0	1	0	0	1	0
:	2	1	1	0	0	1	0	0
:	2	2	1	1	1	1	1	1

- β_3 still estimates the 'difference-in-differences' parameter.

III. EXTENSIONS

What about multiple treated groups?

- Easy to rewrite our earlier model for multiple groups **treated at the same time.**
- 3 units and 3 time periods.
- Groups 1 and 3 implement policy at **T2**.
- **g2** and **g3** are dummies for group 2 and 3
- **t2** and **t3** are respective time dummies for periods 2 and 3.

y	group	time	policy	g2	g3	t2	t3
:	1	1	0	0	0	0	0
:	1	2	1	0	0	1	0
:	1	3	1	0	0	0	1
:	2	1	0	1	0	0	0
:	2	2	0	1	0	1	0
:	2	3	0	1	0	0	1
:	3	1	0	0	1	0	0
:	3	2	1	0	1	1	0
:	3	3	1	0	1	0	1

Extending the model to multiple groups/times

- The regression model with group and time fixed effects would now look something like this (where β_5 is the DD estimate where policy=1):

$$Y_{gt} = \beta_0 + \beta_1 g2 + \beta_2 g3 + \beta_3 t2 + \beta_4 t3 + \beta_5 p_{gt} + \varepsilon_{st}$$

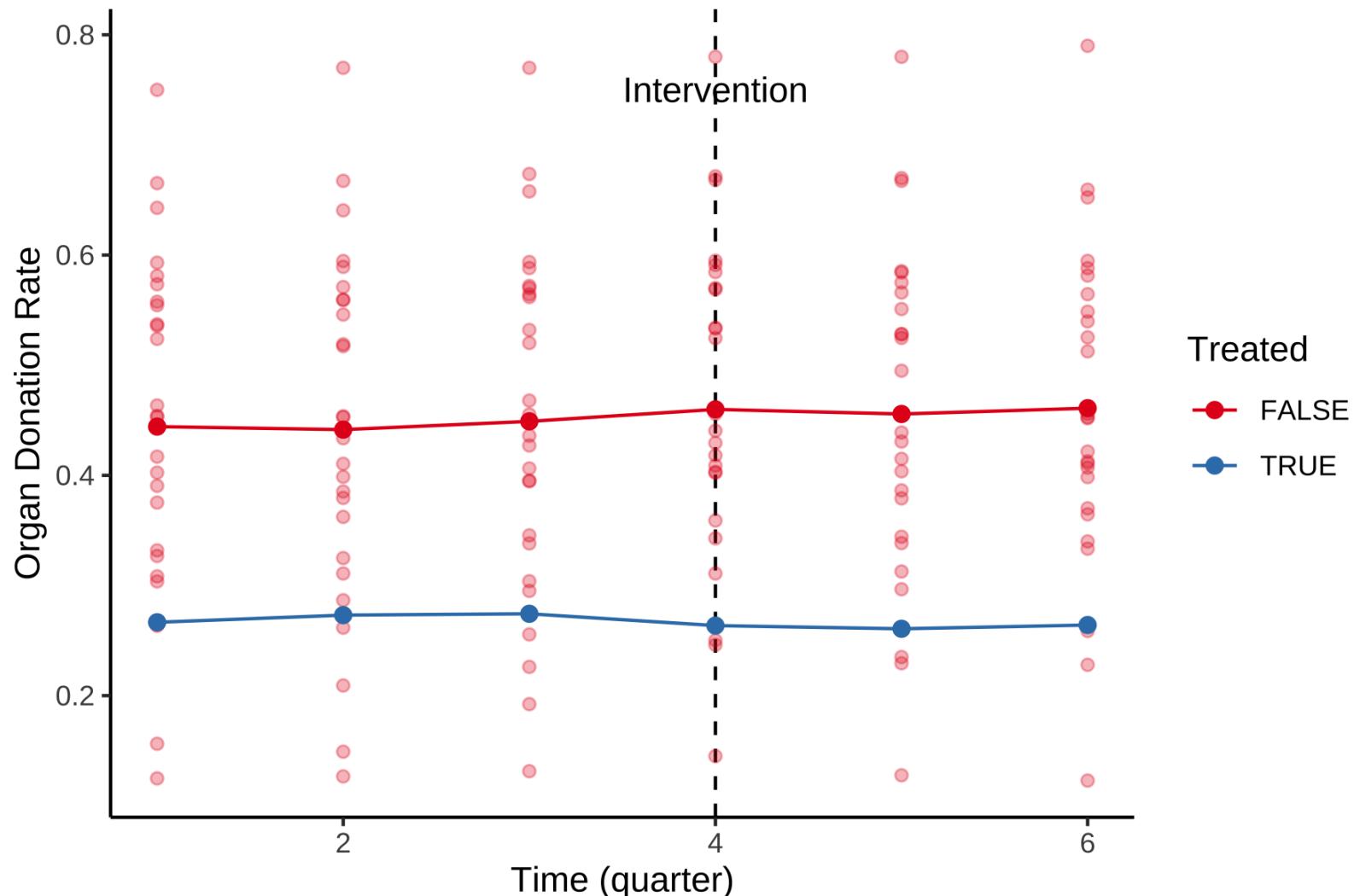
- Reference categories (for interpreting β_0) are group 1 ($g1$) and time 1 ($t1$).
- More generally, you could write the basic equation with multiple group (γ_g) and time (τ_t) fixed effects as:

$$Y_{gt} = \alpha + \gamma_g + \tau_t + \delta^{DD} p_{gt} + \varepsilon_{st}$$

where δ^{DD} is the difference-in-differences estimate for groups treated at time t.

Example

- CA introduced 'opt-out' organ donation policy in July 2011.
- How did this impact organ donation rates?



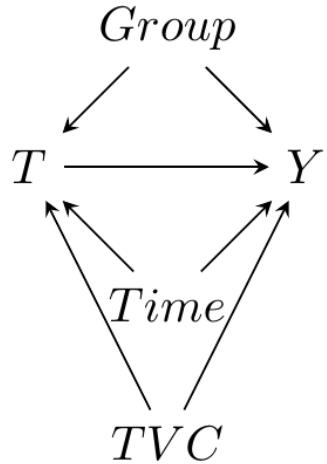
Example

- Pre-post in CA underestimates the effect.
- Treatment comparison in post-period reflects CA's low base rate.
- DD suggests policy reduced organ donation rate.

	Pre-post	Post: T vs. C	DD
Policy	-0.009 (0.003) [-0.016, -0.001]	-0.196 (0.030) [-0.257, -0.135]	-0.022 (0.006) [-0.035, -0.010]
Num.Obs.	6	81	162
Std.Errors	IID	by: State	by: State
FE: Quarter		X	X
FE: State			X

These data are in the `causaldata` package (`organ_donations`) and the DD specification uses the `fixest` package (`feols(Rate ~ Treated | State + Quarter)`)

Extending the basic 2x2 DD



- Note that our basic regression model assumes the only time-varying factor is the policy:
$$Y_{gt} = \alpha + \gamma_g + \tau_t + \delta^{DD} p_{gt} + \varepsilon_{gt}$$
- What if there are confounders of the decision to change the policy?
- That is, we may have omitted important factors that:
 - differ by treatment status.
 - affect the outcome.
 - **are time-varying**, but not affected by the treatment (*TVC*).

The literature on covariates is evolving rapidly. See [Caetano, Callaway, Payne, and Rodrigues \(2022\)](#) for more details.

Extending the basic 2x2 DD: adding time-varying covariates

- E.g., suppose the policy is a soft drink tax and the outcome calories consumed (linear).
- We might worry that *changes in* the density of fast food restaurants could be a common cause of both. We can account for any measured time-varying confounders:

$$Y_{gt} = \alpha + \gamma_g + \tau_t + \delta^{DD} p_{gt} + \zeta Z_{gt} + \varepsilon_{gt}$$

- where ζZ_{gt} is a vector of other controls at the cluster level.
- Important especially if you think other policies may have been implemented simultaneously with treatment.
- Now, conditional on FEs and ζZ_{gt} , we assume that the timing of the change in policy is as good as random.

Assessing heterogeneity

- DD design can also easily be extended to assess differential impact by social group.
- Often of policy interest to assess whether program impacts may differ by:
 - gender.
 - socioeconomic position.
 - disadvantage.
 - levels of the outcome.
- Easy to extend the model to include product term between treatment and "group" of interest.
 - Remember that the treatment varies by time and cluster!
 - So must include all two-way interaction terms with fixed effects.

Evaluating impact on inequalities

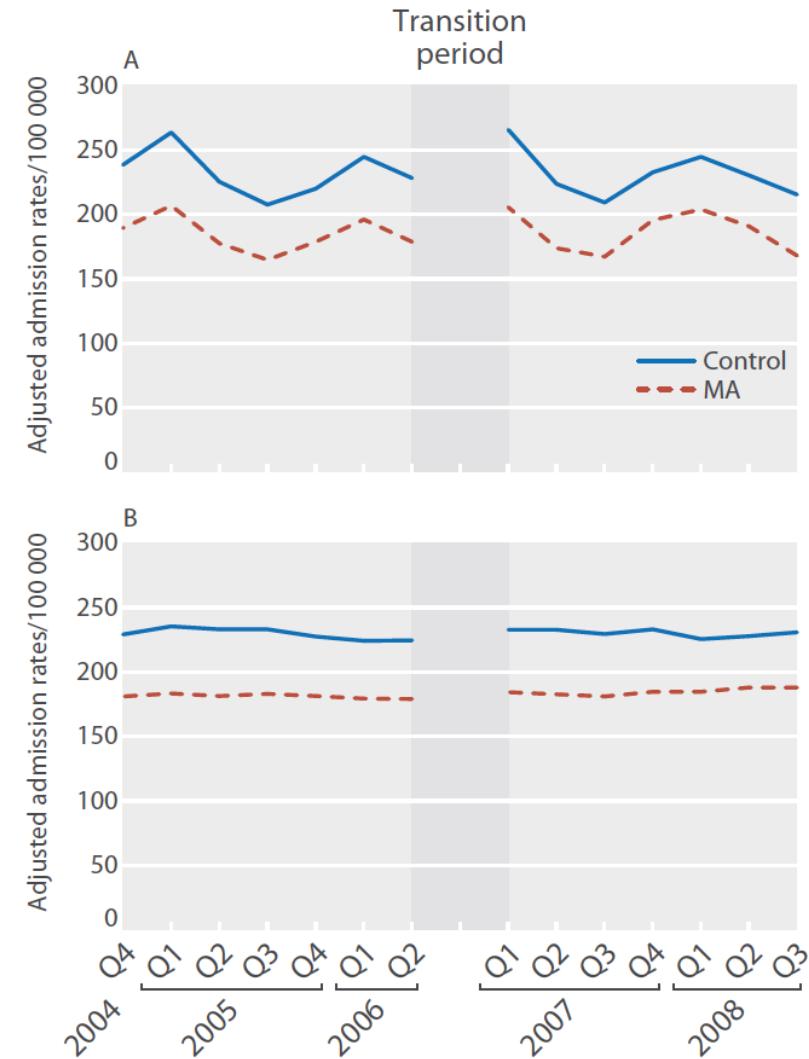
Effect of Massachusetts healthcare reform on racial and ethnic disparities in admissions to hospital for ambulatory care sensitive conditions: retrospective analysis of hospital episode statistics

Danny McCormick,¹ Amresh D Hanchate,^{2,3} Karen E Lasser,³ Meredith G Manze,³ Mengyun Lin,³ Chieh Chu,³ Nancy R Kressin^{2,3}

- Evaluated impact of MA reform on inequalities in hospital admissions.
- Compared MA to nearby states: NY, NJ, PA.
- Intervention "worked": % uninsured halved (12% to 6%) from 2004-06 to 2008-09.

We want credible counterfactuals

- Strong visual evidence that pre-intervention trends similar in treated and control groups.
- Adds credibility to assumption that post-intervention trends **would have been similar** in the absence of the intervention.



Little evidence of differential impact of health reform on racial/ethnic differences in hospital admissions:

Table 3 | Changes in rates of preventable hospital admissions per 100 000 residents/year in Massachusetts and control states (NY, NJ, PA) before (1 October 2004–30 June 2006) and after (1 January 2008–30 September 2009) healthcare reform according to race and ethnicity

ASCS measures	Massachusetts			Control states			Differences in differences estimates		Adjusted estimated % change (95% CI)†
	Before	After	% change	Before	After	% change	Unadjusted	Adjusted (95% CI)*	
Overall composite									
White	667	647	-3.0	716	680	-5.1	2.1	2.1 (-0.8 to 5.0)	Ref
Black	1713	1744	1.8	2188	2240	2.4	-0.6	-0.5 (-6.0 to 5.3)	-1.9 (-8.5 to 5.1)
Hispanic	1258	1203	-4.4	1126	1024	-9.1	4.7	1.6 (-3.9 to 5.5)	2.0 (-7.5 to 12.4)
Acute composite									
White	285	263	-7.5	277	262	-5.6	-1.9	-1.8 (-5.2 to 1.7)	Ref
Black	496	470	-5.3	482	476	-1.2	-4.0	-4.0 (-12.2 to 5.1)	-1.4 (-12.7 to 11.4)
Hispanic	393	362	-7.8	297	276	-7.3	-0.5	-1.2 (-9.9 to 8.3)	2.0 (-10.3 to 15.7)
Chronic composite									
White	383	384	0.3	440	419	-4.7	5.0	5.0 (1.6 to 8.6)	Ref
Black	1217	1274	4.7	1706	1764	3.4	1.3	1.3 (-4.9 to 7.9)	-3.1 (-9.4 to 3.7)
Hispanic	865	840	-2.8	829	748	-9.7	6.9	2.9 (-3.4 to 9.5)	-0.7 (-9.6 to 12.2)

*Adjusted difference in differences estimates and 95% CI obtained from Poisson regression models adjusted for sex, age, race/ethnicity, county income level, county unemployment rate, quarter, and Health Professions Shortage Area designation.

†For change in racial/ethnic disparities in MA v controls. Expresses change in disparities after reform between black and white people and between Hispanic and white people in ACSC (preventable hospitalization) rates after adjustment for changes in control states.

Extensions to non-binary treatments

DD design can also handle treatments/policies/exposures that are not dichotomous.

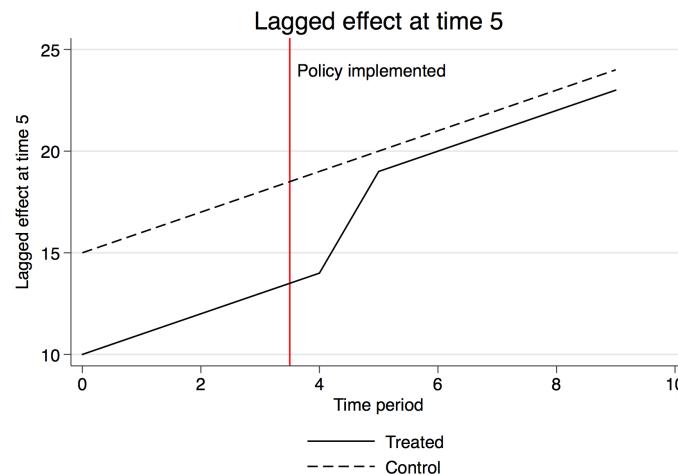
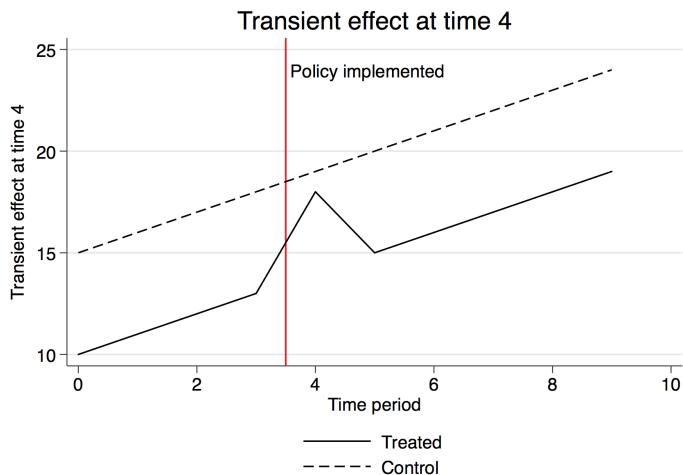
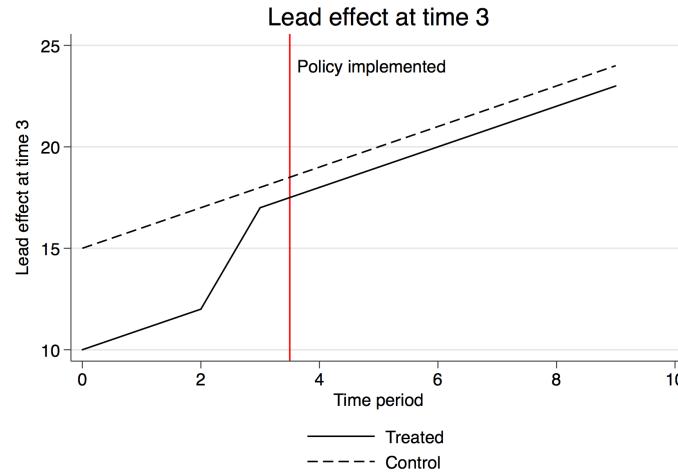
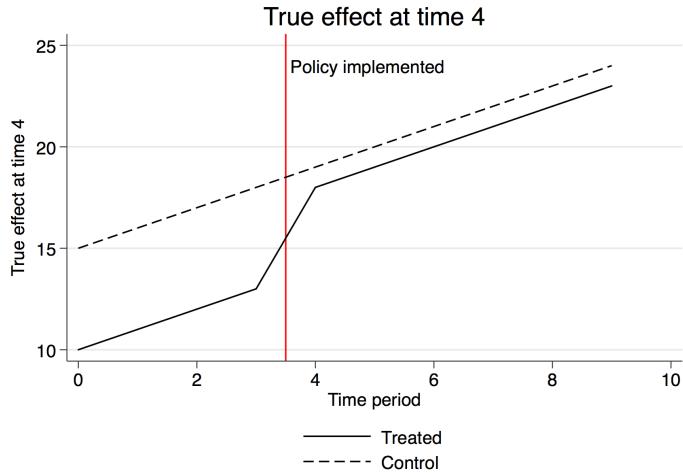
Examples:

- Changes in minimum wage levels (varying "treatment" intensity)
 - Estimate varying levels of increase (\$2 increase vs. \$4)
- "Sin" taxes (e.g., alcohol or cigarettes).
 - differential changes in excise taxes (smaller vs. larger).
- "Weaker" vs. "Stronger" policies
 - texting while driving (primary vs. secondary offense)
 - thresholds for blood alcohol limits (0.15 vs. 0.10 vs. 0.08).

Dynamic Effects

- Basic DD estimates the average ATT over the post-intervention period.
- May average over important variations in how the treatment evolves over time.
- Was the impact immediate? Transient? Sustained over time?
- Can extend the basic model to allow for heterogeneity over time.

Hypothetical dynamic treatment effect scenarios



Basic TWFE applied to 4 scenarios

$$Y_{gt} = \alpha + \gamma_g + \tau_t + \delta^{DD} p_{gt} + \varepsilon_{gt}$$

	(1) True effect	(2) True effect	(3) Transient	(4) Lagged	(5) Lead
Treated group?					
Yes	-5.0	-5.0	-5.3	-4.7	-3.8
Time (years)					
Time period	1.0		0.9	1.1	1.1
Policy					
Yes	4.0	4.0	1.1	2.9	2.7
Time period					
Time period=1		1.0			
Time period=2			2.0		
... etc.					
Time period=9		9.0			
Constant	15.0	15.0	15.4	14.6	14.7
Observations	20	20	20	20	20

What is the problem?

- Challenges with traditional DD model.
- All 4 scenarios demonstrate some "effect" of the policy.
- Transient effect is much smaller in magnitude.
- Lead and lagged effects both suggest positive effects of the policy.
- Basic TWFE model cannot distinguish between these scenarios.

What to do?

- Allow the impact of the treatment to vary over time.
- This allows for *dynamic* effects.
- Center the time variable, then interact the treatment with each time period.
- Can also detect *pre-intervention* effects that suggests concerns about parallel trends.

Extended TWFE applied to 4 scenarios

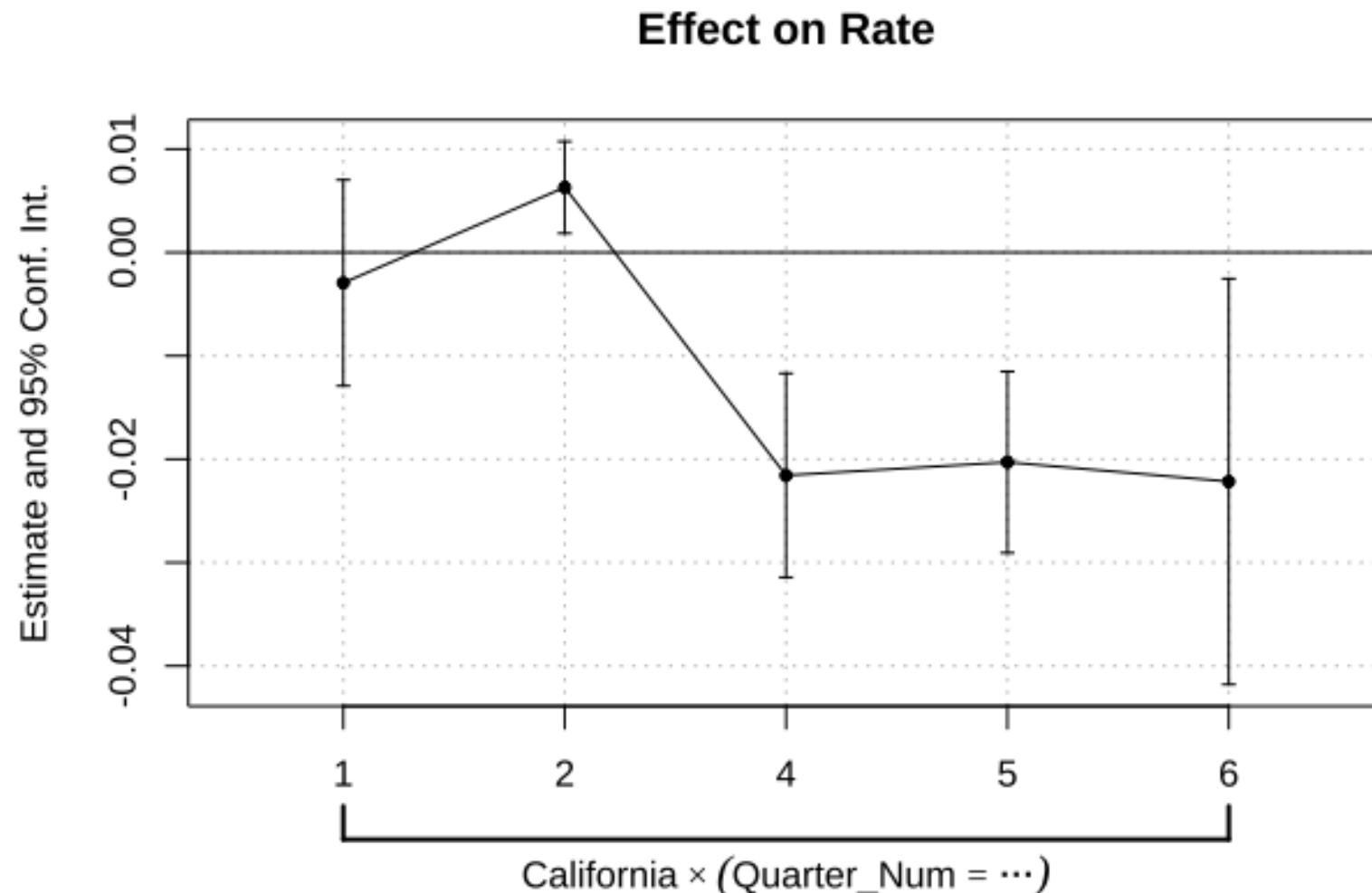
- Allowing for dynamic effects captures heterogeneity.
- Can also allow you to 'see' non-parallel pre-intervention trends.

$$Y_{gt} = \alpha + \gamma_g + \tau_t + \delta_{gt+2} + \delta_{gt+1} + \delta_{gt} + \delta_{gt-1} + \delta_{gt-2} + \delta_{gt-3} + \varepsilon_{gt}$$

	(1) True	(2) Transient	(3) Lagged	(4) Lead
Treated group?				
Yes	-5.0	-5.0	-5.0	-5.0
Time (years)				
Time period	1.0	1.0	1.0	1.0
Leads and lags				
2yr lead	0.0	0.0	0.0	0.0
1yr lead	0.0	0.0	0.0	4.0
actual yr	4.0	4.0	0.0	4.0
1yr lag	4.0	0.0	4.0	4.0
2yr lag	4.0	0.0	4.0	4.0
3+yr lag	4.0	0.0	4.0	4.0
Constant	15.0	15.0	15.0	15.0
Observations	20	20	20	20

CA organ donation example

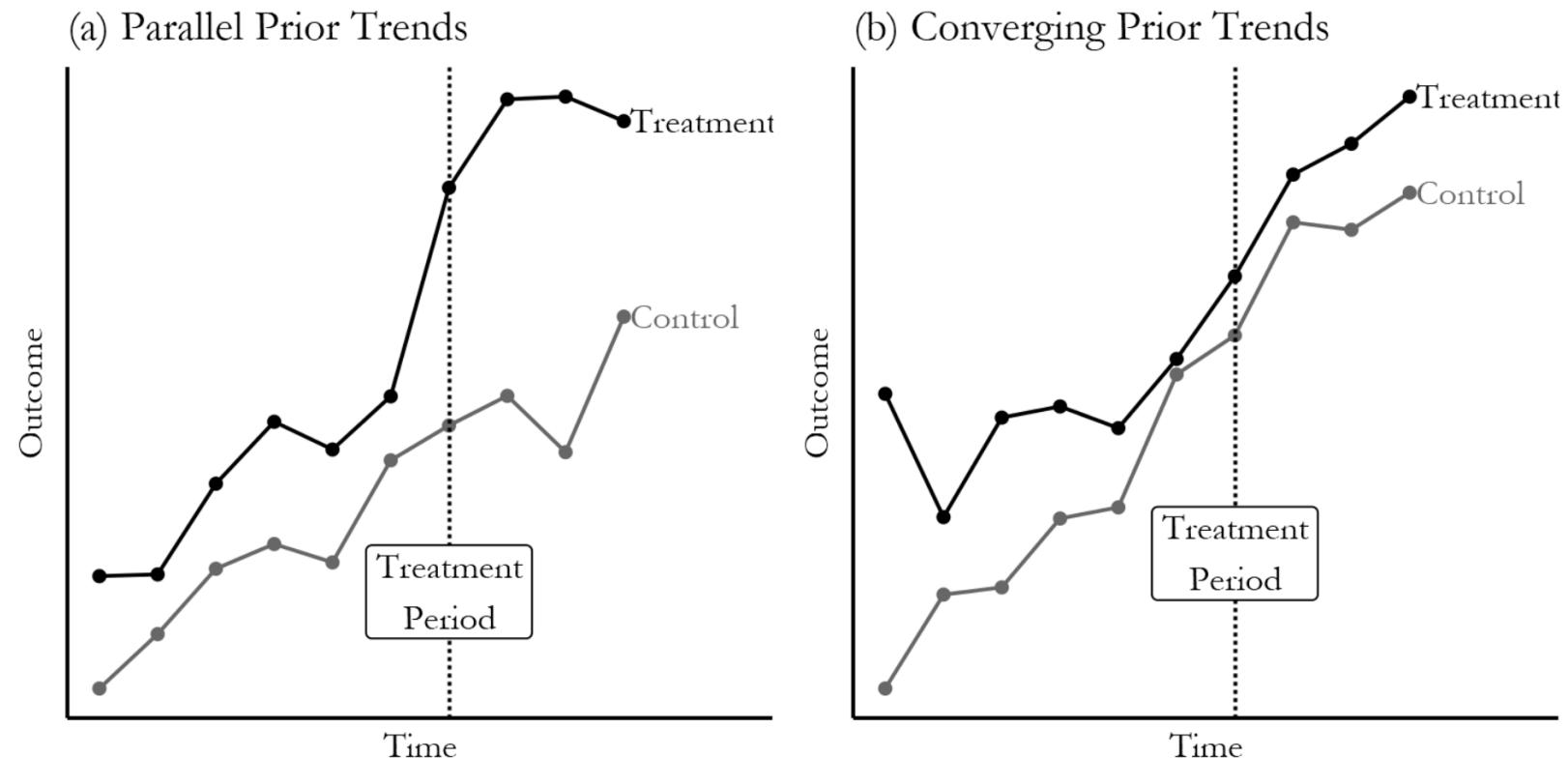
- Impact generally begins post-treatment and remains.
- What about that pre-intervention difference?



IV. ROBUSTNESS CHECKS

Checking parallel trends

- Unverifiable assumption involving the counterfactual.
- Plots of prior trends are helpful.



Source: <https://theeffectbook.net/ch-DifferenceinDifference.html>

A note about scale

- Parallel trends is an assumption about how the *size* of the gap between treated and control would have evolved in the absence of treatment.
- Will depend on functional form for outcome specification (e.g., linear vs. logit).
- Do you care about absolute or relative effects?

Absolute scale (Y)

Area	Before	After	Difference
Treated	135	100	-35
Control	80	60	-20
T - C	55	40	-15

Relative scale ($\ln(Y)$)

Area	Before	After	Difference
Treated	4.9	4.6	0.30
Control	4.4	4.1	0.29
T - C	0.52	0.51	0.01

Evaluating Prior Trends

- Can test whether *pre-intervention* trends differ between treated and control groups.
- Using only pre-intervention data, run:

$$Y_{gt} = \alpha + \gamma_g + \tau_t + \beta_{gt} Group * Time$$

- Tests of β here can be instructive about prior trends.
- Caveats regarding sample sizes and substantive interpretations of any differences.
 - Can find 'significant' but tiny differences with large sample size.
 - Important differences but imprecise so not 'significant'.
 - Mostly consistent but 1 period different?

See recent work by [Rambachan and Roth \(2019\)](#) and [Roth and Sant'Anna \(2020\)](#).

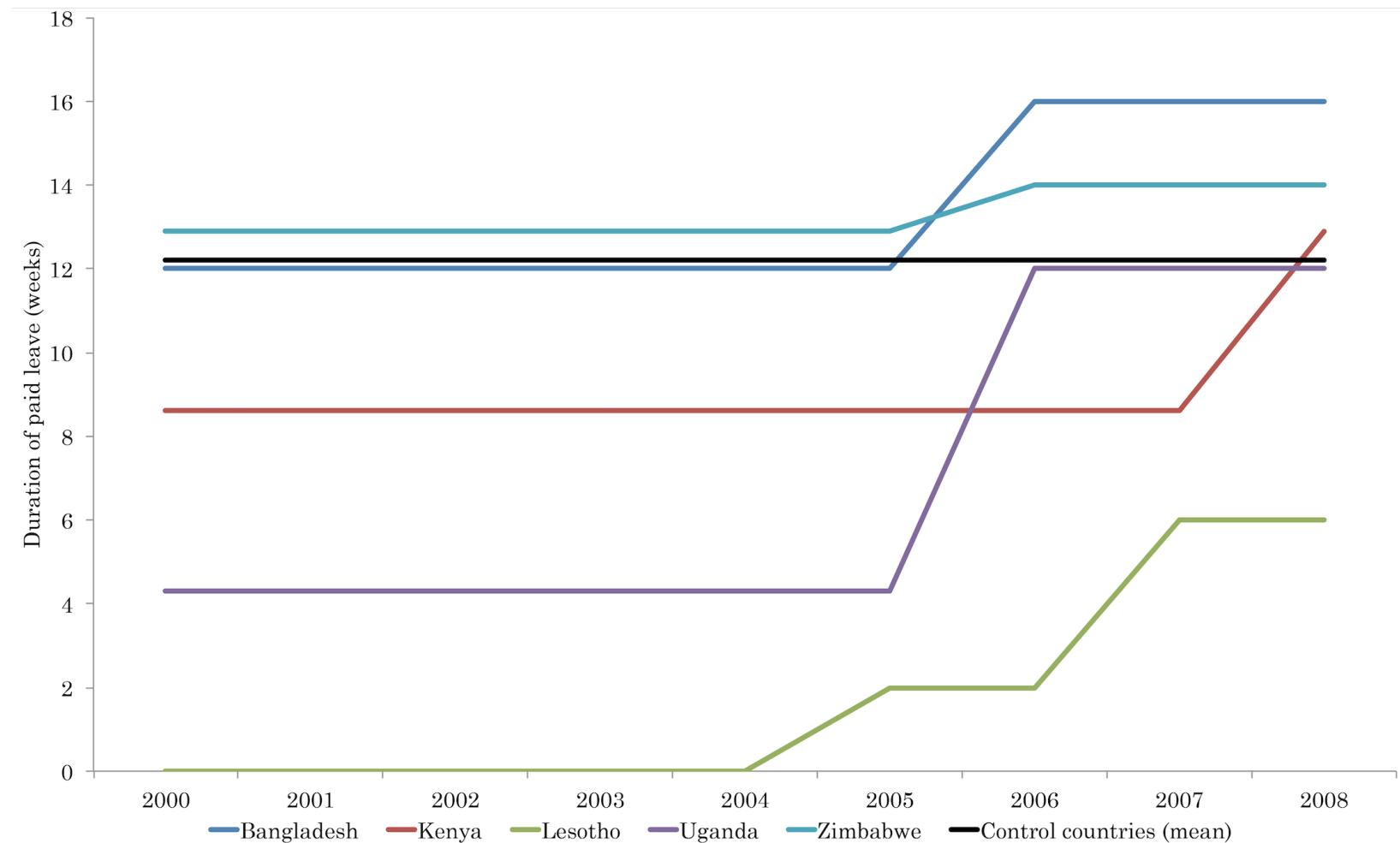
Difference-in-difference-in-differences (yes, this exists.)

- A difference-in-difference-in-differences (DDD) model allows us to study the effect of treatment on different groups.
- If we are concerned that our estimated treatment effect might be spurious, a common robustness test is to introduce a comparison group that should not be affected by the treatment.
- For example, if a policy is designed to affect teenagers we can use a DD model that takes advantage of policy variation across states, and then use a DDD model to study how the policy has affected younger versus older individuals.
- Assuming no true spillover effects.

V. STAGGERED TREATMENTS

What about staggered treatments?

- Different groups adopt treatments at different times.
- Creates many 2x2 DDs.



Recall the setup for DD regression with multiple treated groups

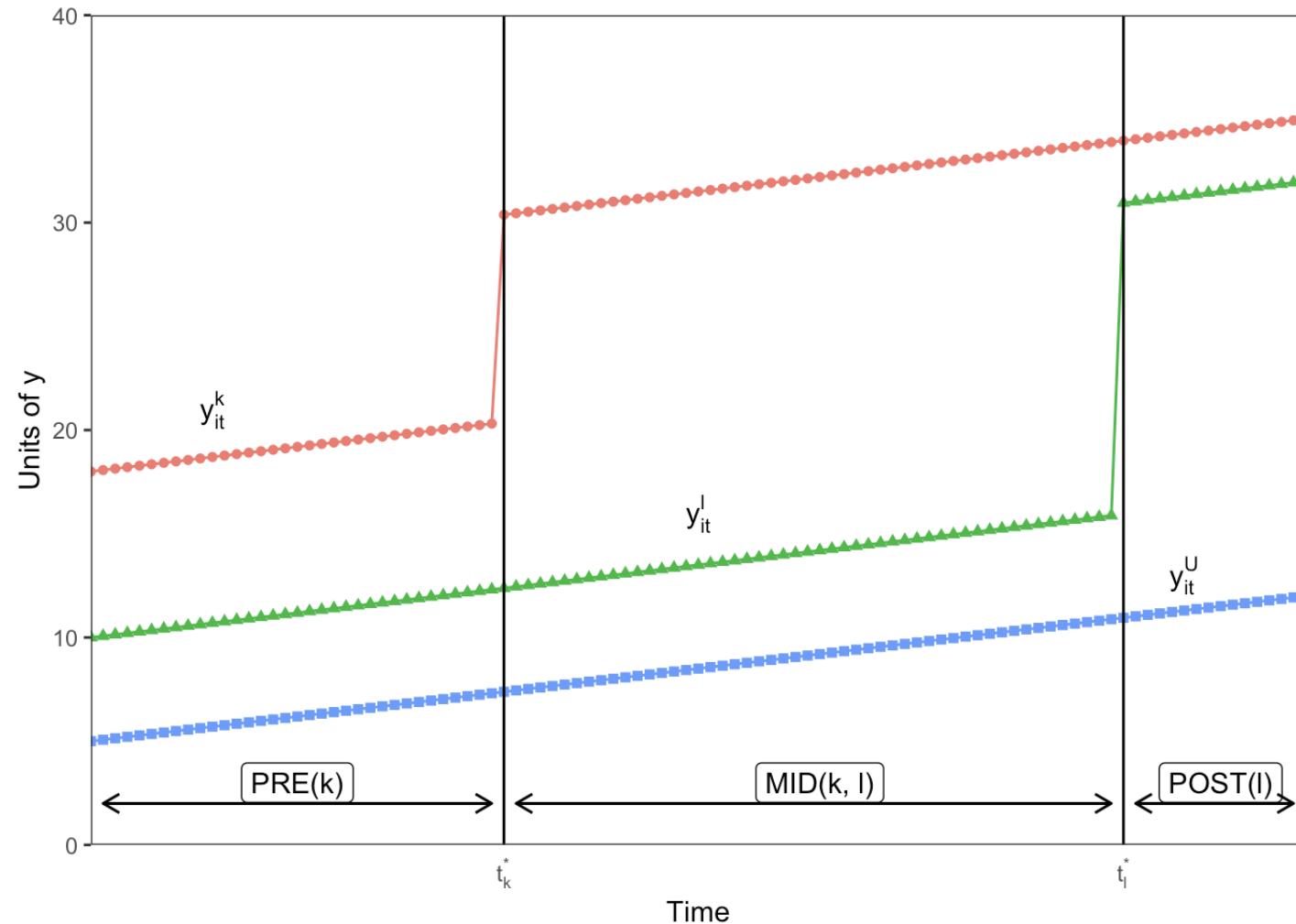
We have different units that are exposed to our intervention at different times. We often use OLS (or LPM) to fit:

$$y_{gt} = \alpha + \gamma_g + \tau_t + \beta^{DD} p_{gt} + \epsilon_{gt}$$

where

- y_{it} is the outcome for unit i at time t .
- γ_g are group-specific fixed effects.
- τ_t are fixed effects for each time period.
- p_{gt} is a time-varying treatment indicator.
- β^{DD} is the difference-in-differences estimate.

1. Early-adopters (k) vs. never treated (U)
2. Later-adopters (l) vs. never treated (U).
3. Early (k) vs. later (l) adopters.
- 4. Later (l) vs. earlier (k) adopters.**



Graph from <https://andrewcbaker.netlify.app/2019/09/25/difference-in-differences-methodology/>

What is the problem?

- Using earlier treated groups as controls only 'works' if the treatment effects are:
 - Homogeneous across groups at a given time; and
 - Homogeneous over time (no dynamic effects).
- This adds any changes in treatment effects in the early group, which get **subtracted from the DD estimate**.
- Can lead to β^{DD} that is a poor summary of group-specific effects if there is heterogeneity.

What are potential solutions?

- All basically involve **not allowing** early treated groups to serve as controls later.

Callaway and Sant'Anna (2021)

Use non-parametric group-time ATTs (+ covariates).

Abraham and Sun (2021)

Use saturated fixed effects to ensure that prior treated units are not used as controls

Cengiz, Dube, Lindner, and Zipperer (2019)

Create state-event-specific panel datasets and calculate event-specific estimates using separate regressions for each state-event.

Many new papers on this. See recent papers by [Callaway and Sant'Anna \(2021\)](#), [Goodman-Bacon \(2021\)](#), [Cengiz, Dube, Lindner, and Zipperer \(2019\)](#) and [Sun and Abraham \(2021\)](#)

Key Takeaways

- DD is a common and powerful design to estimate policy and program impacts.
- Compares *changes* in outcomes in a treated group to a control group.
- Controls for time-invariant unobserved group factors *and* common trends in outcomes.
- Requires good qualitative knowledge about *why* the treated group became treated.
- Core assumption is parallel trends, unverifiable but not impossible to investigate.
- Many extensions possible (inequalities, multiple groups and periods, non-binary treatments).

References

- Benmarhnia, T. et al. (2019). "A rose by any other name still needs to be identified (with plausible assumptions)". In: *Int J Epidemiol* 48.6, pp. 2061-2062. DOI: [10.1093/ije/dyz049](https://doi.org/10.1093/ije/dyz049).
- Caetano, C. et al. (2022). "Difference in Differences with Time-Varying Covariates". In: *arXiv*. DOI: [10.48550/ARXIV.2202.02903](https://doi.org/10.48550/ARXIV.2202.02903). URL: <https://arxiv.org/abs/2202.02903>.
- Callaway, B. et al. (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200-230.
- Cengiz, D. et al. (2019). "The effect of minimum wages on low-wage jobs". In: *The Quarterly Journal of Economics* 134.3, pp. 1405-1454.
- Freedman, D. A. (1991). "Statistical models and shoe leather". In: *Sociological methodology* 21.2, pp. 291-313.
- Gertler, P. J. et al. (2011). *Impact evaluation in practice*. World Bank Publications.

References (cont)

- [Goodman-Bacon, A.](#) (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254-277.
- [Kessler, J. B. et al.](#) (2014). *Don't take 'no' for an answer: An experiment with actual organ donor registrations*. National Bureau of Economic Research.
- [Lopez Bernal, J. et al.](#) (2019). "Difference in difference, controlled interrupted time series and synthetic controls". In: *Int J Epidemiol* 48.6, pp. 2062-2063. DOI: [10.1093/ije/dyz050](https://doi.org/10.1093/ije/dyz050).
- [McCormick, D. et al.](#) (2015). "Effect of Massachusetts healthcare reform on racial and ethnic disparities in admissions to hospital for ambulatory care sensitive conditions: retrospective analysis of hospital episode statistics". In: *BMJ* 350, p. h1480.
- [Raifman, J. et al.](#) (2020). "State handgun purchase age minimums in the US and adolescent suicide rates: regression discontinuity and difference-in-differences analyses". In: *BMJ* 370, p. m2436. DOI: [10.1136/bmj.m2436](https://doi.org/10.1136/bmj.m2436).

References (cont)

- Rambachan, A. et al. (2019). "An honest approach to parallel trends". In: *Unpublished manuscript, Harvard University*.
- Roth, J. et al. (2020). "When is parallel trends sensitive to functional form?" In: *arXiv preprint arXiv:2010.04814*.
- Snow, J. et al. (1965). *Snow on cholera: being a reprint of two papers*. New York: Hafner Pub. Co.
- Sun, L. et al. (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175-199.
- Venkataramani, A. S. et al. (2017). "Health consequences of the US Deferred Action for Childhood Arrivals (DACA) immigration programme: a quasi-experimental study". In: *Lancet Public Health* 2.4, pp. e175-e181. DOI: [10.1016/S2468-2667\(17\)30047-6](https://doi.org/10.1016/S2468-2667(17)30047-6).