USAID
FROM THE AMERICAN PEOPLE

Rescuing impact measurements

Differential treatment timing and duration in d-i-d estimation

Session Objectives

By the end of this session, participants will

- Understand the basic setup of quasi-experimental difference-in-differences
- Understand how d-i-d generalizes to multiple time periods and groups, and how complications can arise from this generalization
- Recognize adjustments that can be made to handle additional complexity
- Appreciate that generating defensible quasi-experimental impact estimates is difficult

Outline of Presentation

- Background
- Problem
- Solutions
- Case study - MISTI
- Final thoughts

Bottom Line Up Front

- The difference-in-differences estimator can generalize to multiple treatment groups and multiple time periods
- However, under certain conditions that we only realized recently, this can introduce bias
- Examine the different groups created by differential timing
- Use event study designs and other newer estimators that account for this bias

- **Background**
- Problem
- Solutions
- Case study - MISTI
- Final thoughts

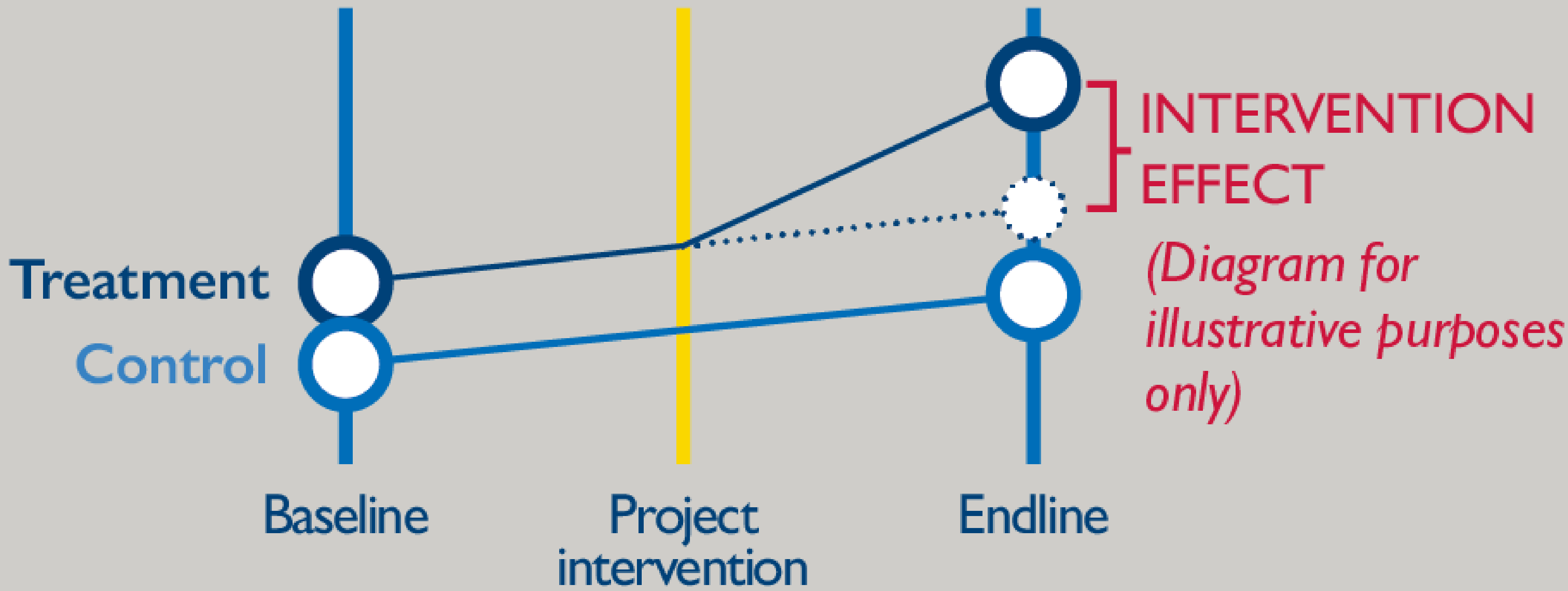
How Do We Do Difference-in-differences?

- Under randomized d-i-d, the pre-treatment measurement is used to improve precision
- Under quasi-experimental d-i-d, we depend on the trend to remove time-invariant sources of bias
- For this to work, we must demonstrate or convincingly argue for parallel trends
- If we justify parallel pre-treatment trends, then we can use the breaks in trends after treatment to estimate the treatment effect

test

\textcolor$redx+2$

Quasi-experimental d-i-d



What Does This Look Like

Group	Baseline	Endline	Difference
Treatment	12	18	6
Comparison	14	16	2
Difference	-2	2	4

Treatment group difference = 18 - 12 = 6

Comparison group difference = 16 - 14 = 2

Treatment difference - Comparison difference = 6 - 2 = 4

Baseline difference - Endline difference = 2 - (-2) = 4

test

Group	Baseline	Endline	Difference
Comparison	14	16	2
Treatment	12	18	6
Difference	-2	2	4

What is the Canonical d-i-d Setup?

where..

is the comparison group at baseline

is the change in comparison group from baseline to endline

is the baseline difference between the treatment and comparison

is the treatment effect, the interaction of treatment and time test

test

Group	Baseline	Endline	Difference
Comparison	14	16	2
Treatment	12	18	6
Difference	-2	2	4

test

Group	Baseline	Endline	Difference
Treatment	12	18	6
Comparison	14	16	2
Difference	-2	2	4

Canonical d-i-d, 2x2

	Pre	Post	Post - Pre
Comparison			
Treatment			
Treatment - Comparison			

Generalizing d-i-d to Many Periods/Groups

When we generalize to multiple time periods and/or groups, we have the two-way fixed effect (TWFE) estimator

where..

are group fixed effects

are time fixed effects

indicates whether group *i* in period *t* is treated

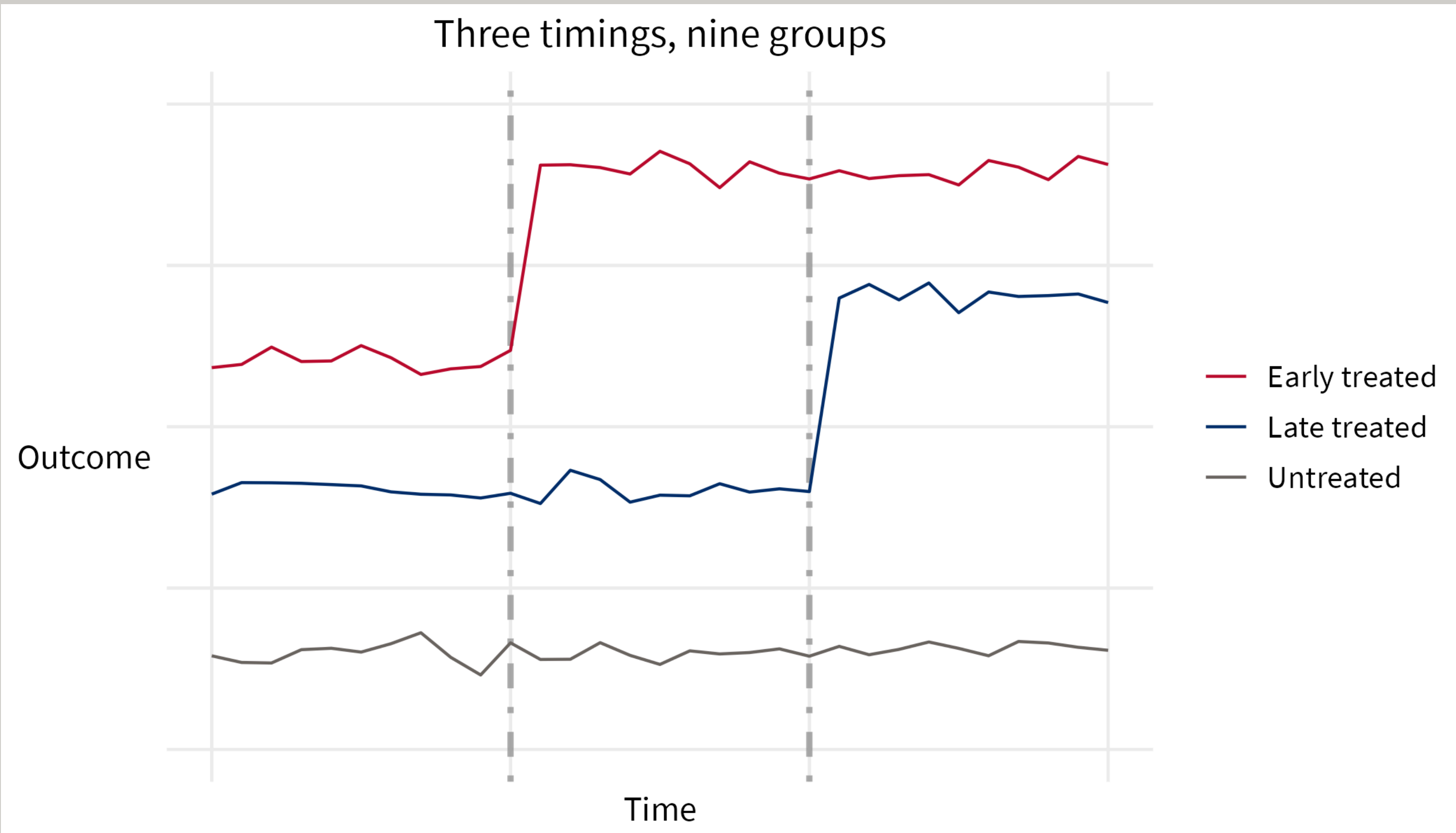
- Background
- **Problem**
- Solutions
- Case study - MISTI
- Final thoughts

But What is Actually Telling Us?

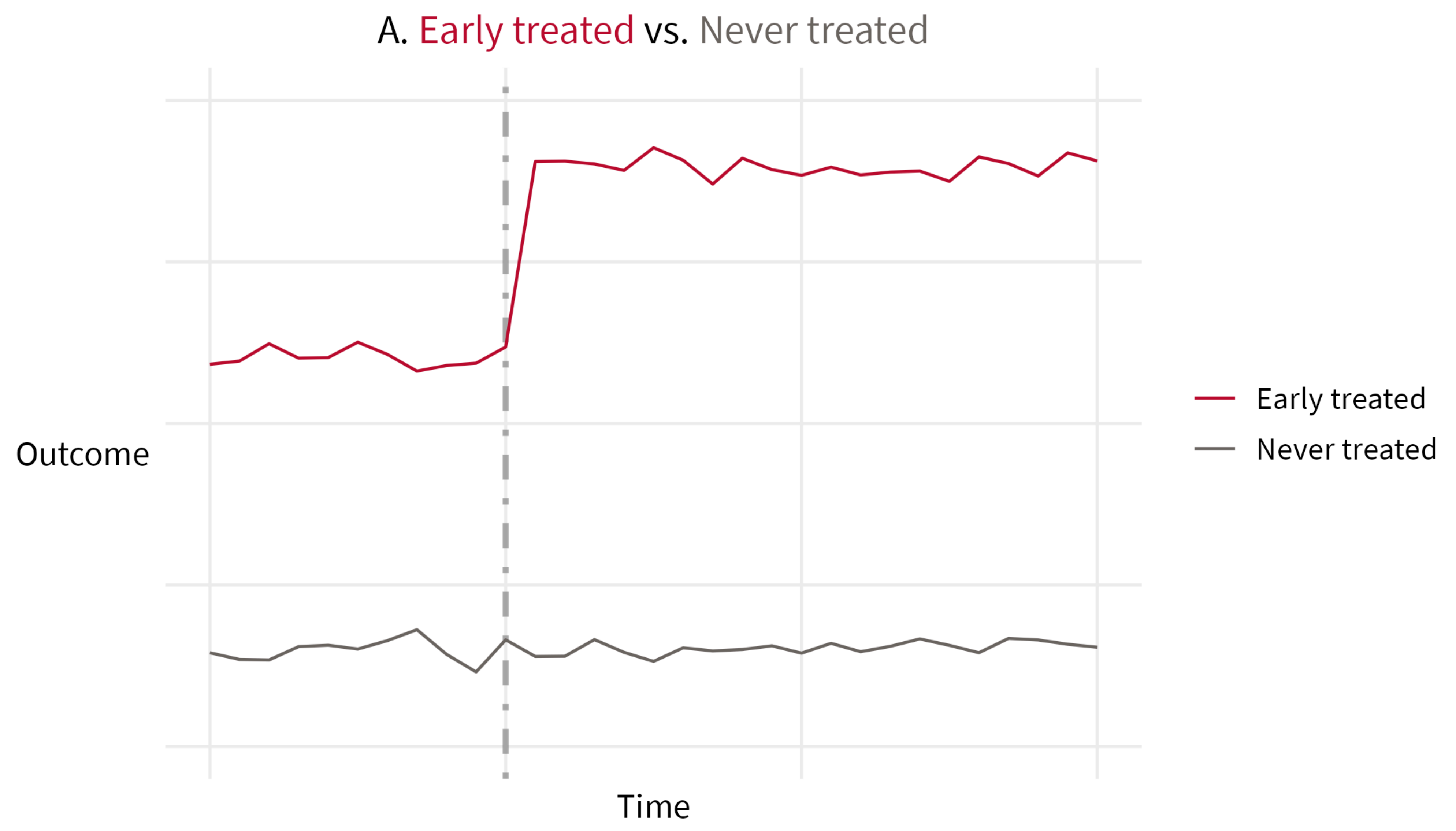
- For the canonical 2x2, we know exactly what we are estimating
- For *i* groups and *t* time periods, we are getting some average of multiple 2x2s
- But how does this work, exactly?

Two Treatment Groups, Early and Late

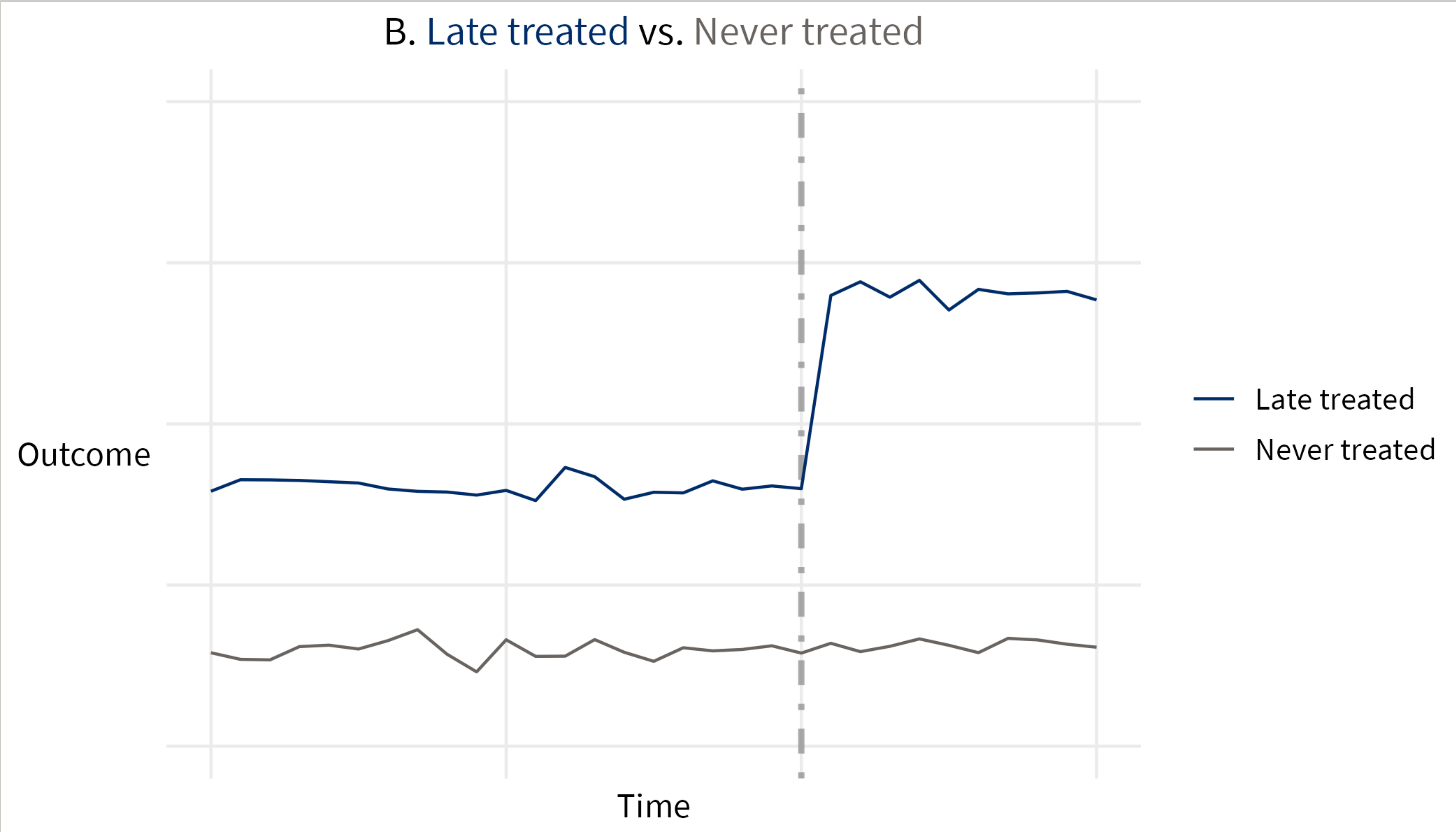
Let's take a single step from two time periods to three, where treatment can be adopted at either t = 2 or t = 3



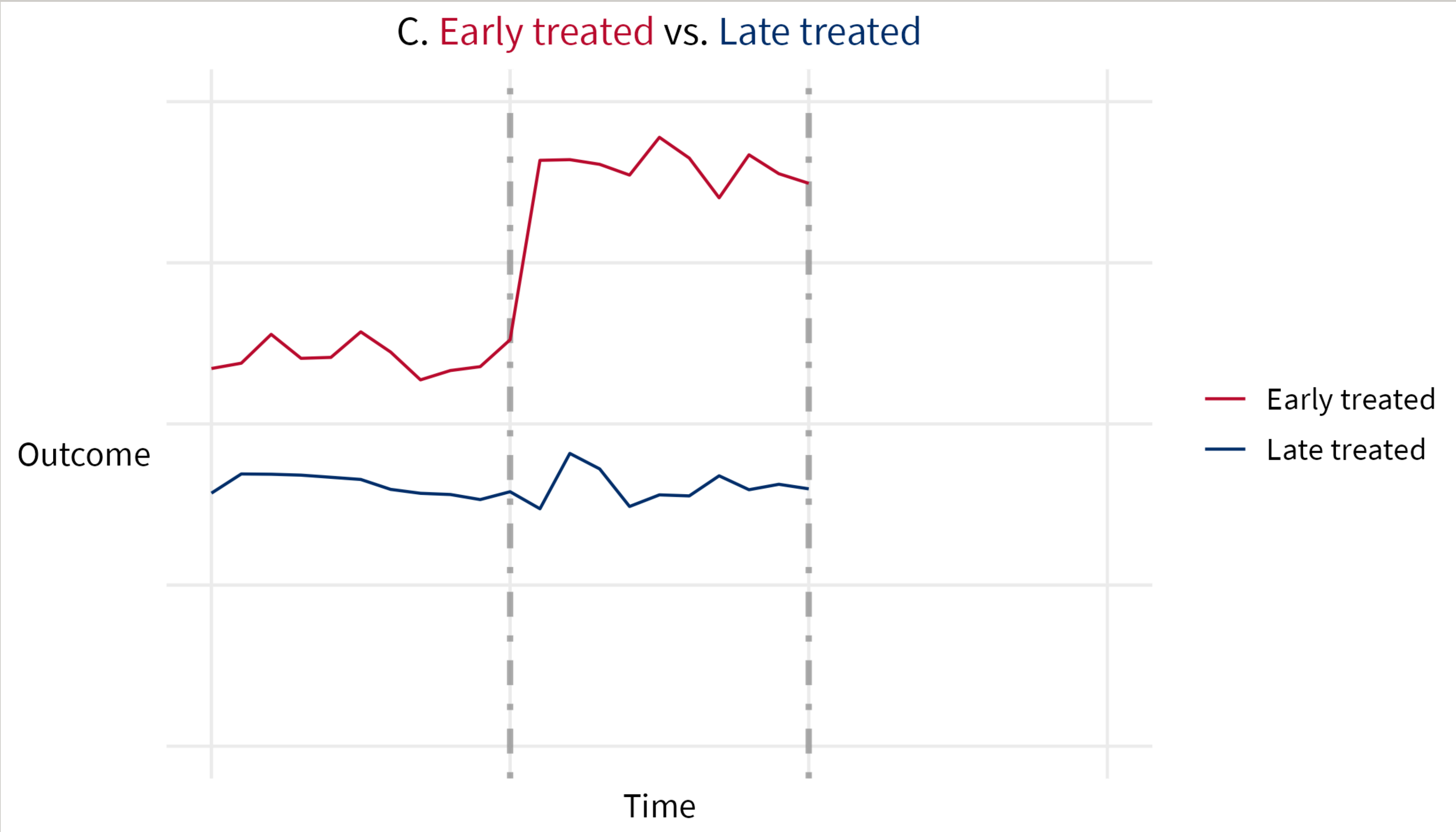
Not a Problem



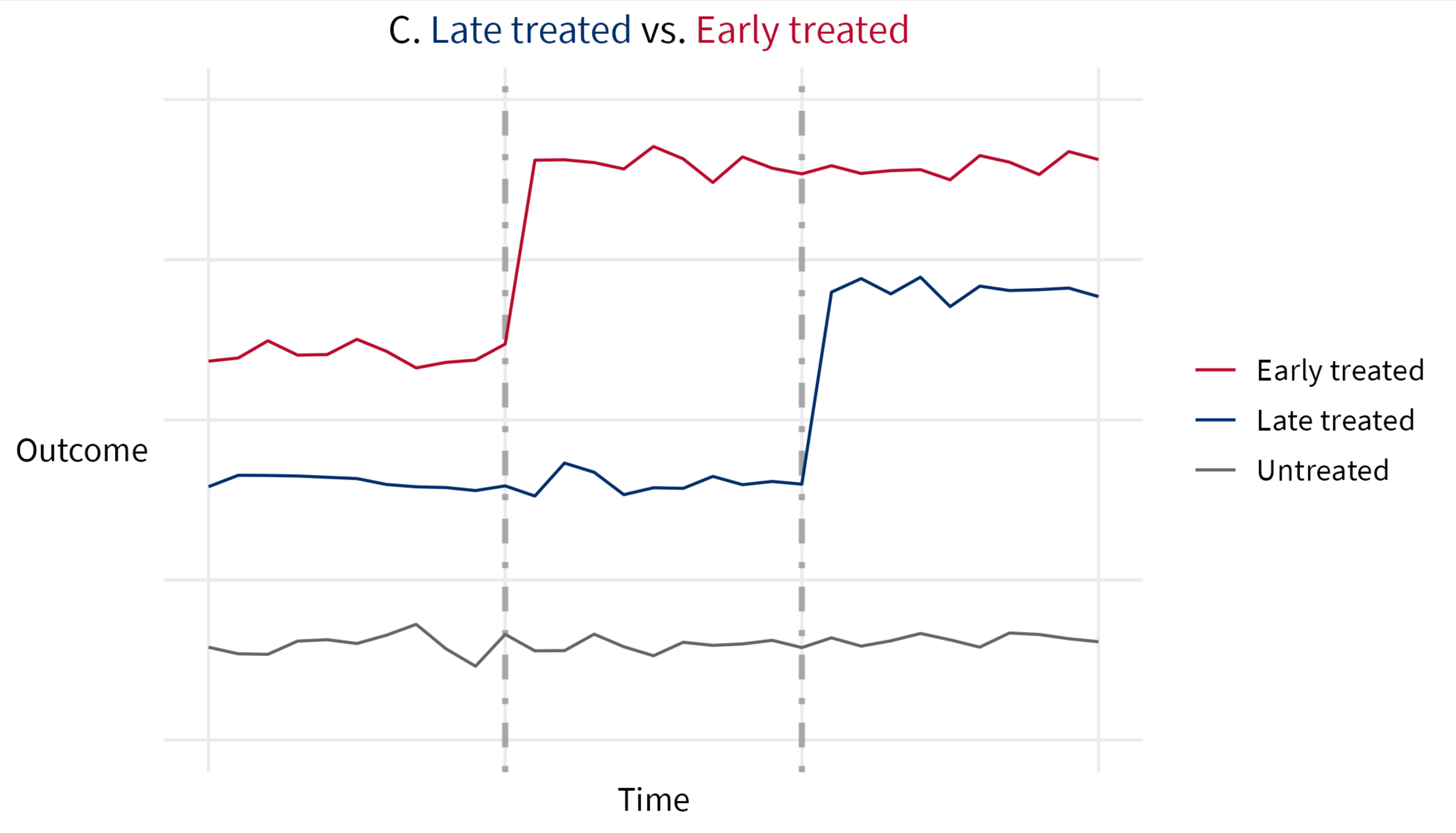
Not a Problem



Not a Problem



COULD BE PROBLEM!



Where Does This Leave Us?

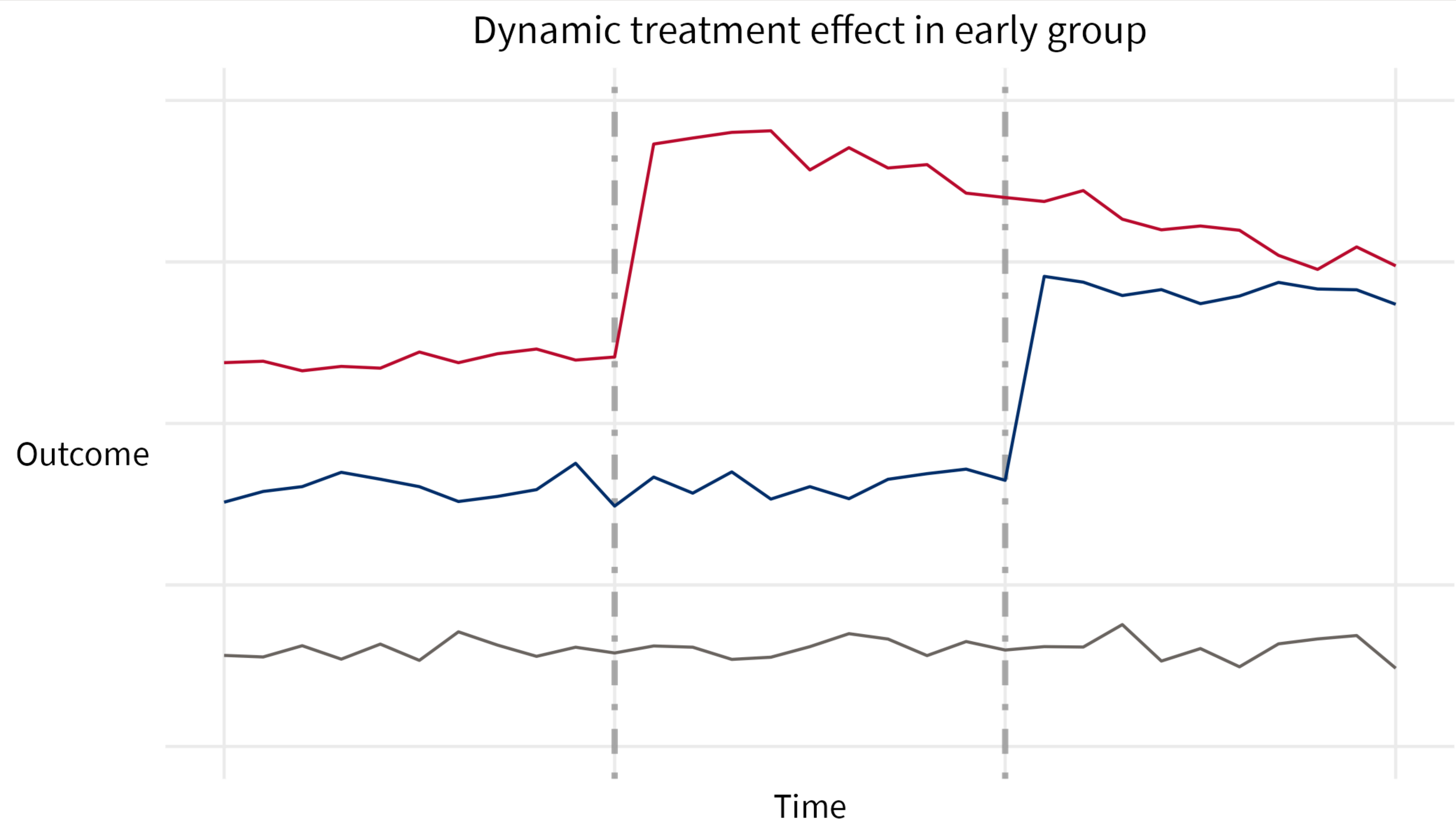
- TWFE treats some data that is under treatment status as comparison!
- Not an issue under constant treatment effect
  - Stable unit treatment value (SUTVA)
  - No variation in treatment effect for any reason

TWFE Fails

But TWFE fails under the following conditions:

- Different treatment groups have different treatment effects
- Treatment effects are dynamic over post-treatment periods
- Heterogeneous treatment effects across sub-groups within a treated group

Definitely a Problem



- Background
- Problem
- **Solutions**
- Case study - MISTI
- Final thoughts

Adjustment: New Estimators

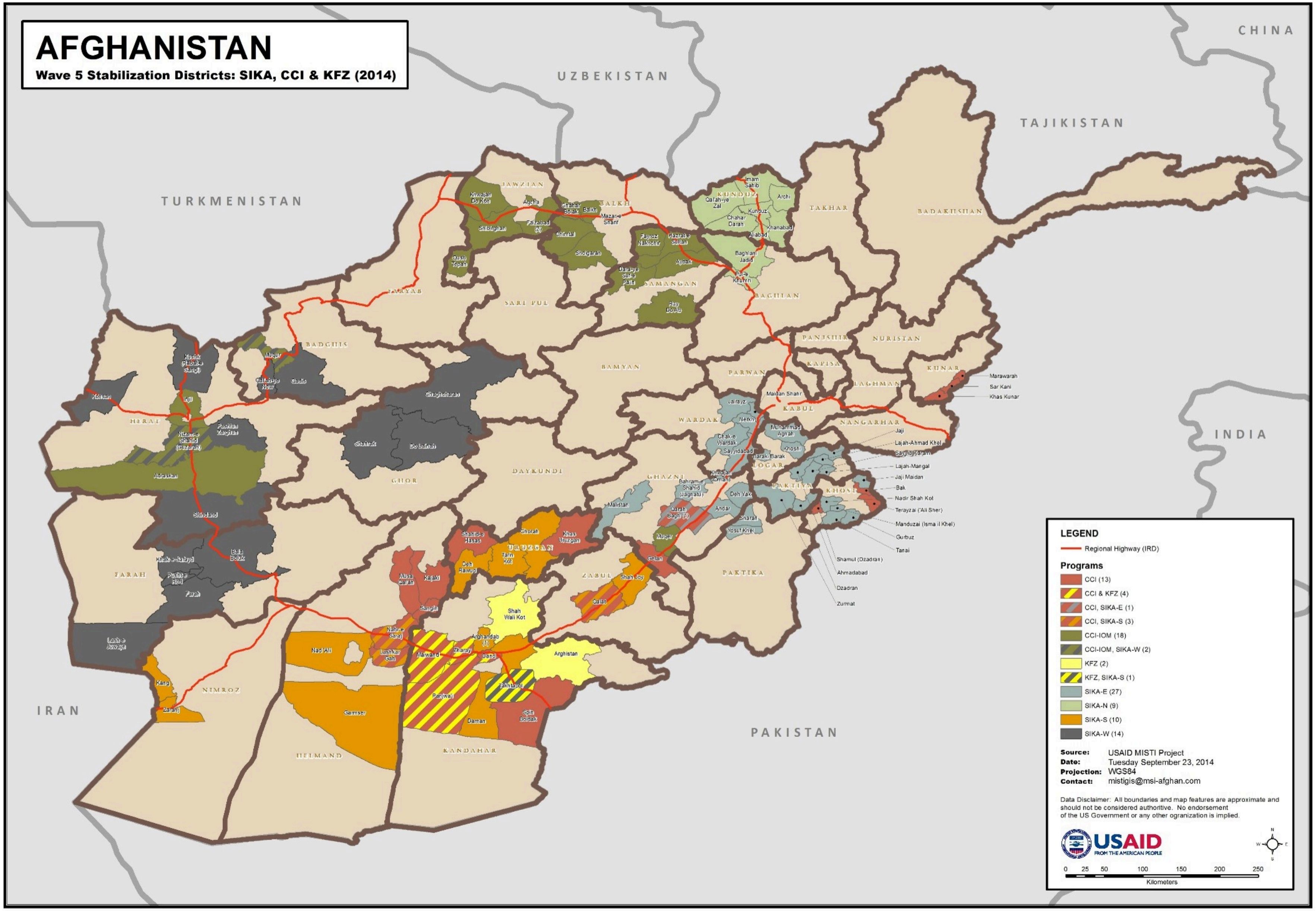
- Background
- Problem
- Solutions
- **Case study - MISTI**
- Final thoughts

Measuring Impact of Stabilization Initiatives (MISTI)

Can small scale, community-driven development activities build local government legitimacy in a kinetic conflict-affected environment?

MISTI

- Village panel survey in five waves, Sep 2012 - Nov 2014
- ~5,000 villages surveyed across 130 districts and 23 provinces
- ~ 30,000 household interviews per wave
- 860 treated villages at any wave (17%)
- 355 villages surveyed in all five waves
- 85 villages treated (24%)


AFGHANISTAN
Wave 5 Stabilization Districts: SIKA, CCI & KFZ (2014)
UZBEKISTAN
TURKMENISTAN
TAJIKISTAN
CHINA
INDIA
PAKISTAN
IRAN
JAWZJAN
BALKH
KUNDUZ
TAKHAR
BADAKHSHAN
FARYAB
SARI PUL
SAMANGAN
BAGHLAN
PANJSHIR
NURISTAN
BADGHIS
BAMYAN
PARWAN
KAPISA
KUNAR
LAGHMAN
KABUL
WARDAK
NANGARHAR
LOGAR
HERAT
GHOR
DAYKUNDI
GHAZNI
PAKTIYA
KHOST
PAKTIKA
URUZGAN
ZABUL
FARAH
NIMROZ
HELMAND
KANDAHAR
Marawarah
Sar Kani
Khas Kunar
Jaji
Lajah-Ahmad Khel
Lajan-Mangal
Jaji Maidan
Bak
Nadir Shah Kot
Terayzai ('Ali Sher)
Manduzai (Isma il Khel)
Gurbuz
Tanai
Shamul (Dzadran)
Ahmadabad
Dzadran
Zurmat
LEGEND
Regional Highway (IRD)
Programs
CCI (13)
CCI & KFZ (4)
CCI, SIKA-E (1)
CCI, SIKA-S (3)
CCI-IOM (18)
CCI-IOM, SIKA-W (2)
KFZ (2)
KFZ, SIKA-S (1)
SIKA-E (27)
SIKA-N (9)
SIKA-S (10)
SIKA-W (14)
Source: USAID MISTI Project
Date: Tuesday September 23, 2014
Projection: WGS84
Contact: mistigis@msi-afghan.com
Data Disclaimer: All boundaries and map features are approximate and should not be considered authoritive. No endorsement of the US Government or any other ogranization is implied.
USAID
FROM THE AMERICAN PEOPLE
0 25 50 100 150 200 250
Kilometers

MISTI Treatment Timings

Wave	Comparison Villages	Treated Villages	Cumulative Treated
1	355	0	0
2	341	14	14
3	322	19	33
4	302	20	53
5	270	32	85

Single-wave Analysis

Final reporting of MISTI relied on a series of single-wave estimations

Measure	Wave 2-4	Wave 2-5	Wave 3-4	Wave 4-5	Wave 3-5
Stability	.031	.043	.003	-.039	-.002

MISTI Two Way Fixed Effects (TWFE)

Term	Estimate	Standard Error	‘t-statistic’	‘p-value’
Treatment	-.0389	.0947	-.411	.681

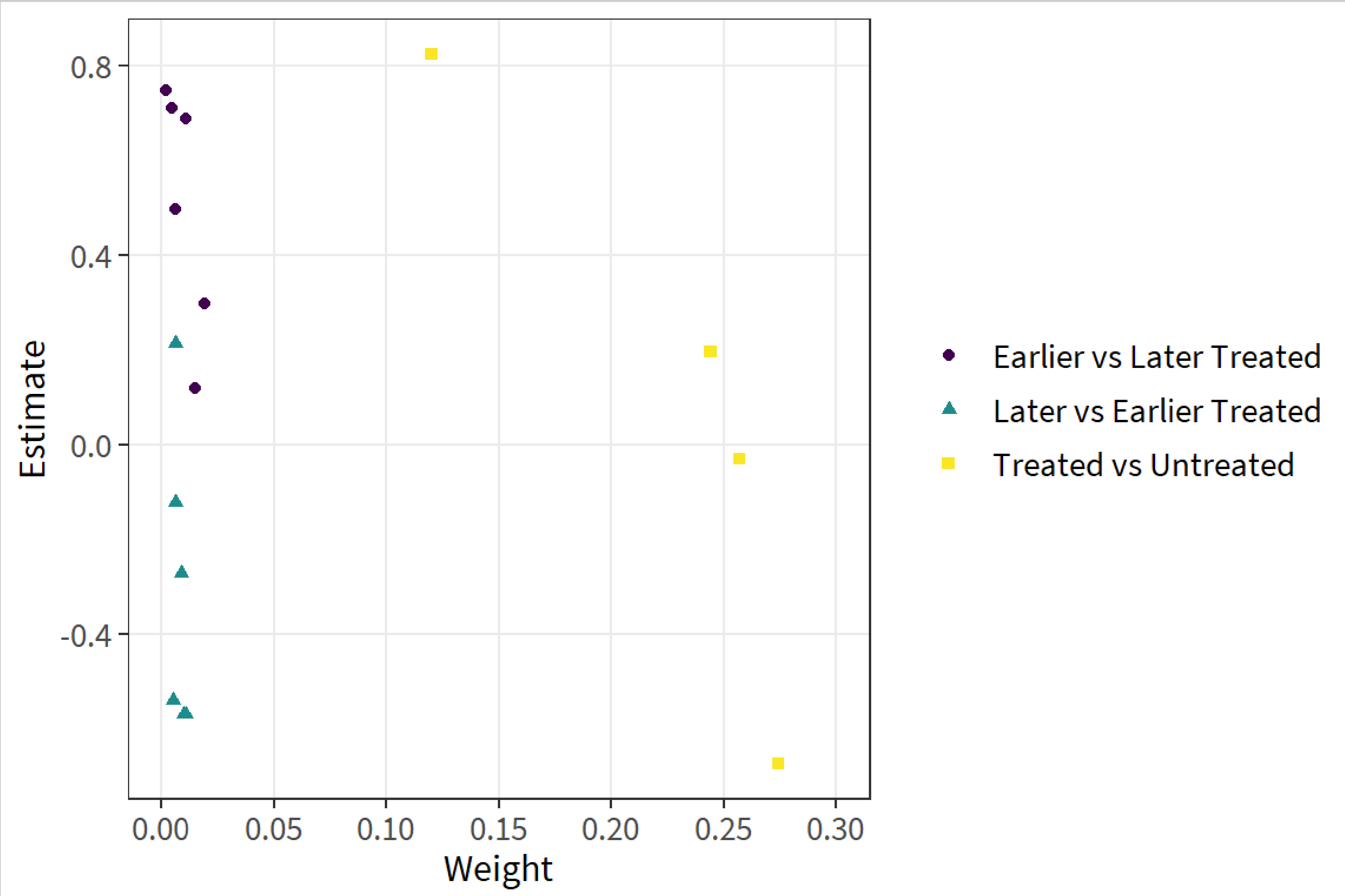
Diagnostic: the Bacon Decomposition

- The Bacon decomposition will take a TWFE model and decompose it into the full array of 2x2 d-i-d cells used to construct the overall estimate
- The decomposition will also calculate the variance-weights used in regression to see which 2x2 cells are powering the overall estimate
- After diagnosing a problem, the researcher can correct for the bias by using the newer estimators

MISTI bacondecomp 2x2 Cells

treated	untreated	estimate	weight	type
4	5	0.1182	0.01524	Earlier vs Later Treated
3	5	0.2976	0.01930	Earlier vs Later Treated
2	5	0.6868	0.01066	Earlier vs Later Treated
5	4	-0.5375	0.00508	Later vs Earlier Treated
3	4	0.4973	0.00603	Earlier vs Later Treated
2	4	0.7093	0.00444	Earlier vs Later Treated
5	99999	-0.6729	0.27424	Treated vs Untreated
4	99999	-0.0291	0.25710	Treated vs Untreated
3	99999	0.1964	0.24425	Treated vs Untreated
2	99999	0.8232	0.11998	Treated vs Untreated
5	3	-0.5686	0.00965	Later vs Earlier Treated
4	3	0.2135	0.00603	Later vs Earlier Treated
2	3	0.7473	0.00211	Earlier vs Later Treated
5	2	-0.5680	0.01066	Later vs Earlier Treated
4	2	-0.2707	0.00889	Later vs Earlier Treated
3	2	-0.1216	0.00633	Later vs Earlier Treated

Plot of 2x2 Cells



MISTI bacondecomp

```
                      type weight avg_est
1 Earlier vs Later Treated 0.0578  0.3911
2 Later vs Earlier Treated 0.0466 -0.3465
3     Treated vs Untreated 0.8956 -0.0506
```

treated	untreated	estimate	weight	type
4	5	0.118	0.0152	Earlier vs Later Treated
3	5	0.298	0.0193	Earlier vs Later Treated
2	5	0.687	0.0107	Earlier vs Later Treated

5	4	-0.537	0.00508	Later vs Earlier Treated
3	4	0.497	0.00603	Earlier vs Later Treated
2	4	0.709	0.00444	Earlier vs Later Treated
5	1e+05	-0.673	0.274	Treated vs Untreated
4	1e+05	-0.0291	0.257	Treated vs Untreated
3	1e+05	0.196	0.244	Treated vs Untreated
2	1e+05	0.823	0.12	Treated vs Untreated
5	3	-0.569	0.00965	Later vs Earlier Treated

4	3	0.214	0.00603	Later vs Earlier Treated
2	3	0.747	0.00211	Earlier vs Later Treated
5	2	-0.568	0.0107	Later vs Earlier Treated
4	2	-0.271	0.00889	Later vs Earlier Treated
3	2	-0.122	0.00633	Later vs Earlier Treated

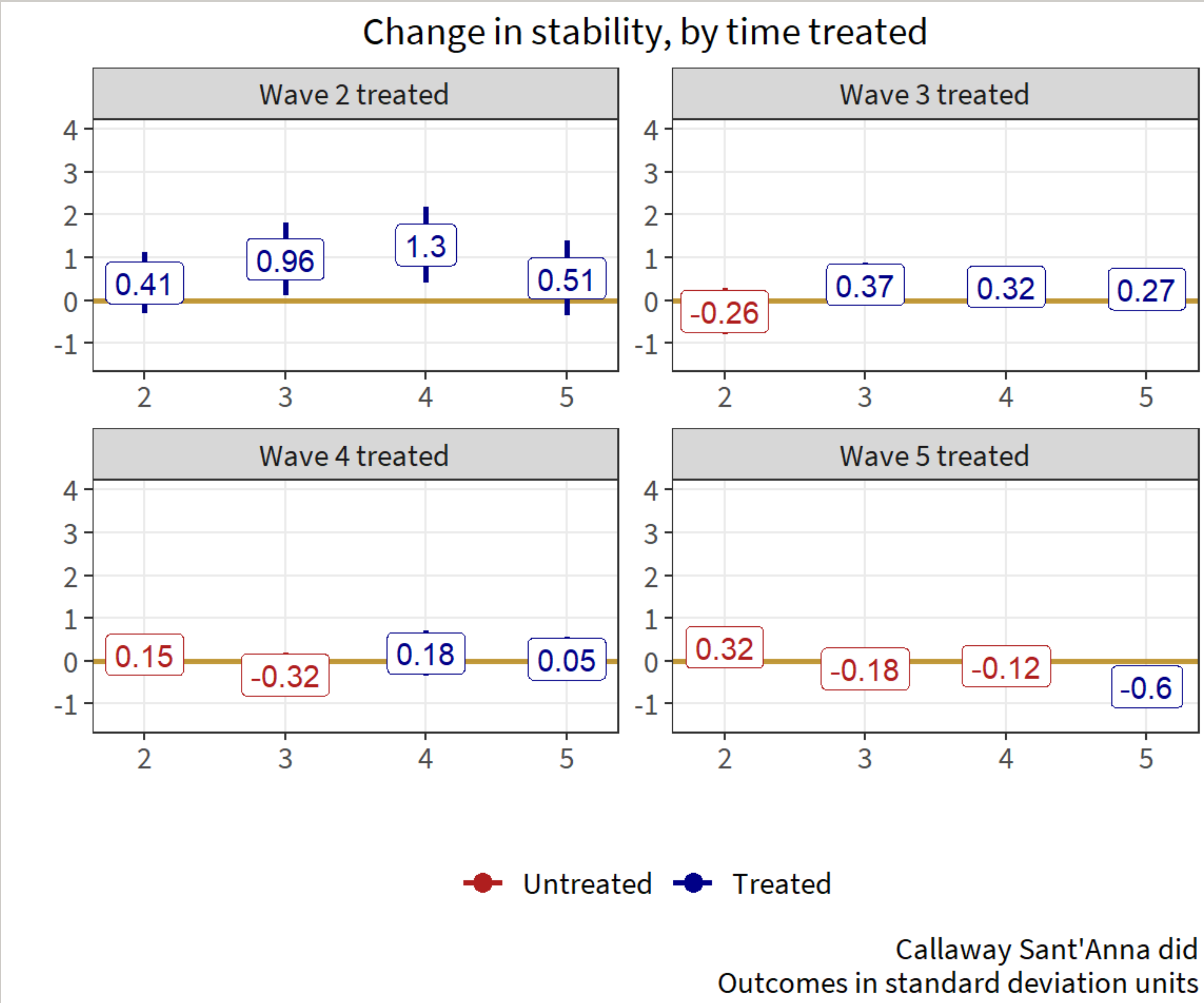
```
[1] -0.0389
```

Callaway and Sant'Anna (2021)

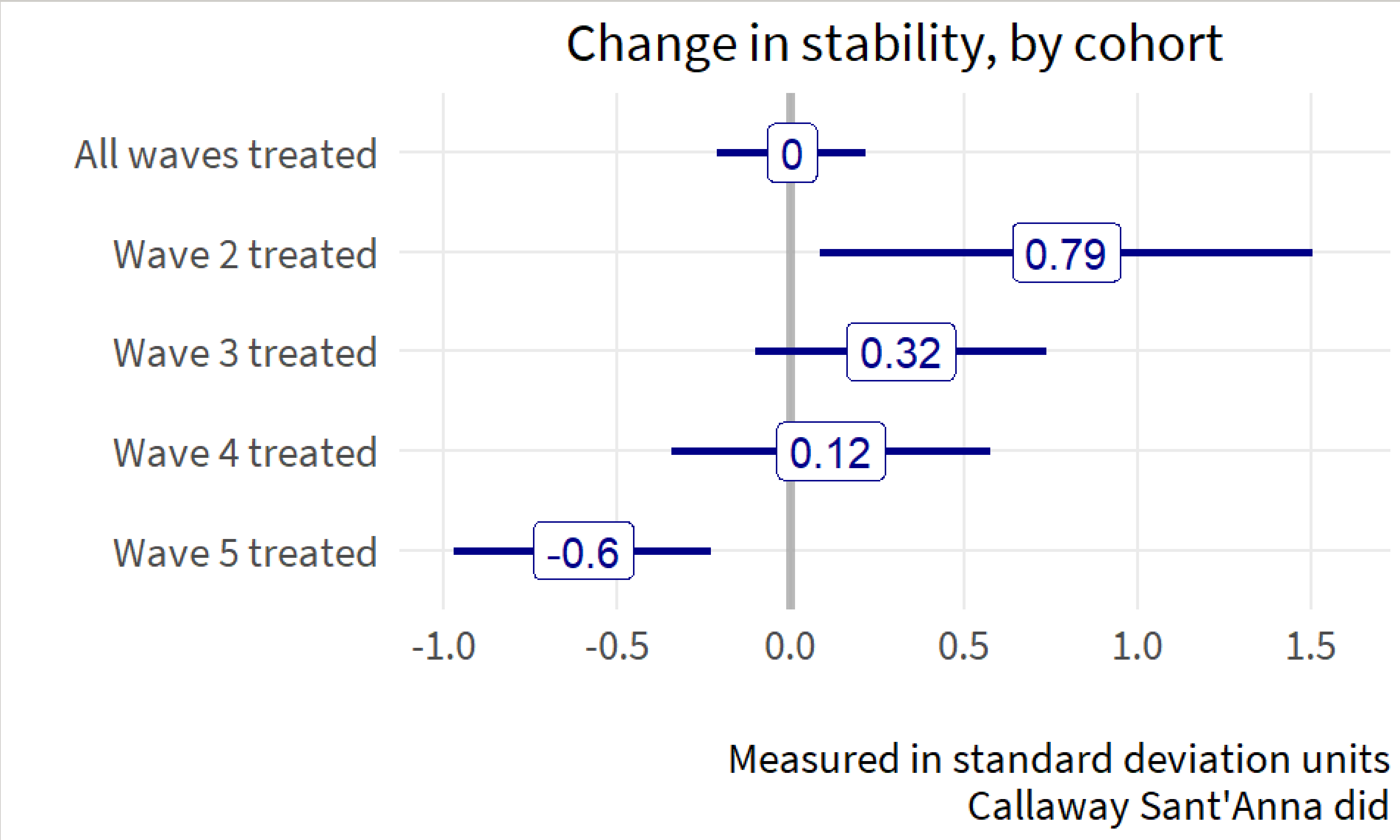
This estimation gives you multiple outcomes

- Treatment by treatment group
- An overall treatment effect
- Overall dynamic effects / event study

Treatment Effects for Each Wave



Overall Effects by Cohort



- Background
- Problem
- Solutions
- Case study - MISTI
- **Final thoughts**

What Have We Learned About MISTI?

- Using the newer econometric estimators, we were able to show dynamic treatment effects across time
- These dynamic effects highlighted initial success of the 2010 surge
- Early success gradually declined until the legitimacy crisis of the 2014 election

What Have We Learned About Evaluation?

- In certain settings, two-way fixed effects estimation is biased in ways that we only recently came to realize
- We have to carefully think through the data generating process (logic modeling) for each individual setting
- As we get more granular data and ask deeper questions, econometric tools are starting to provide better insight into treatment dynamics

What Should We Do?

- For any two-way fixed effects setting, use the Bacon decomposition to diagnose any problems
- Use event study designs to examine dynamic treatment effects
- Re-examine old evaluations!!

Thank you!