Rescuing impact measurements

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

2024-09-06

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

What is the canonical d-i-d setup?

where..

β_0 is the comparison group at baseline

δ_0 is the secular change from baseline to endline, unrelated to treatment

β_1 is the difference between the treatment and comparison groups at baseline, and

δ_1 is the treatment effect, the interaction of treatment and time

Algebraically, δ_0 can be expressed as the difference between the pre/post difference in each of the treatment and comparison groups

$$\delta_1 = (\bar{y}_{POST,TREAT} - \bar{y}_{PRE,TREAT}) - (\bar{y}_{POST,COMPARISON} - \bar{y}_{PRE,COMPARISON})$$

hence, difference-in-differences (d-i-d or DiD or DD)

Canonical d-i-d, 2x2

$$y_{it} = \beta_0 + \delta_{0,t}Post_t + \beta_{1,i}Treat_i + \delta_{1,it}Post_t * Treat_i + \epsilon_{it}$$

	Pre	**Post**	**Post - Pre**
Comparison	β_0	$\beta_0 + \delta_0$	δ_0
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment - Comparison	β_1	$\beta_1 + \delta_1$	δ_1

Generalizing d-i-d to multiple periods or groups

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

$$y_{it} = \alpha_i + \alpha_t + \beta_{it}^{DD} + \epsilon_{it}$$

where..

α_i are group fixed effects

α_t are time fixed effects

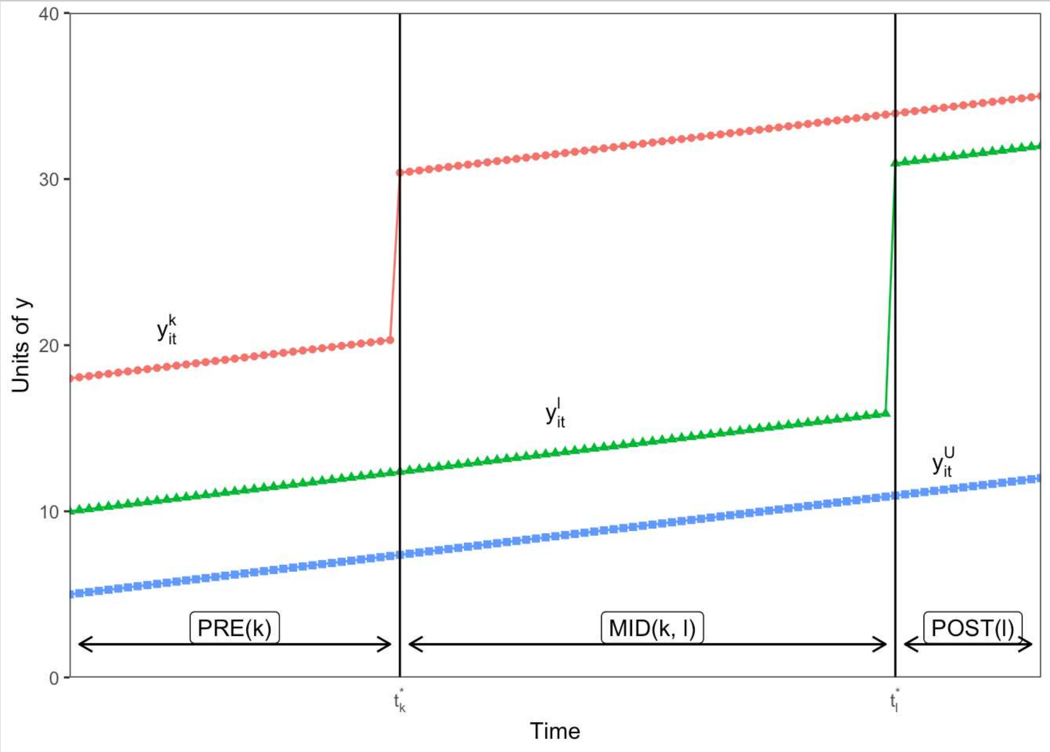
B_{it}^{DD} indicates whether group *i* in period *t* is treated

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

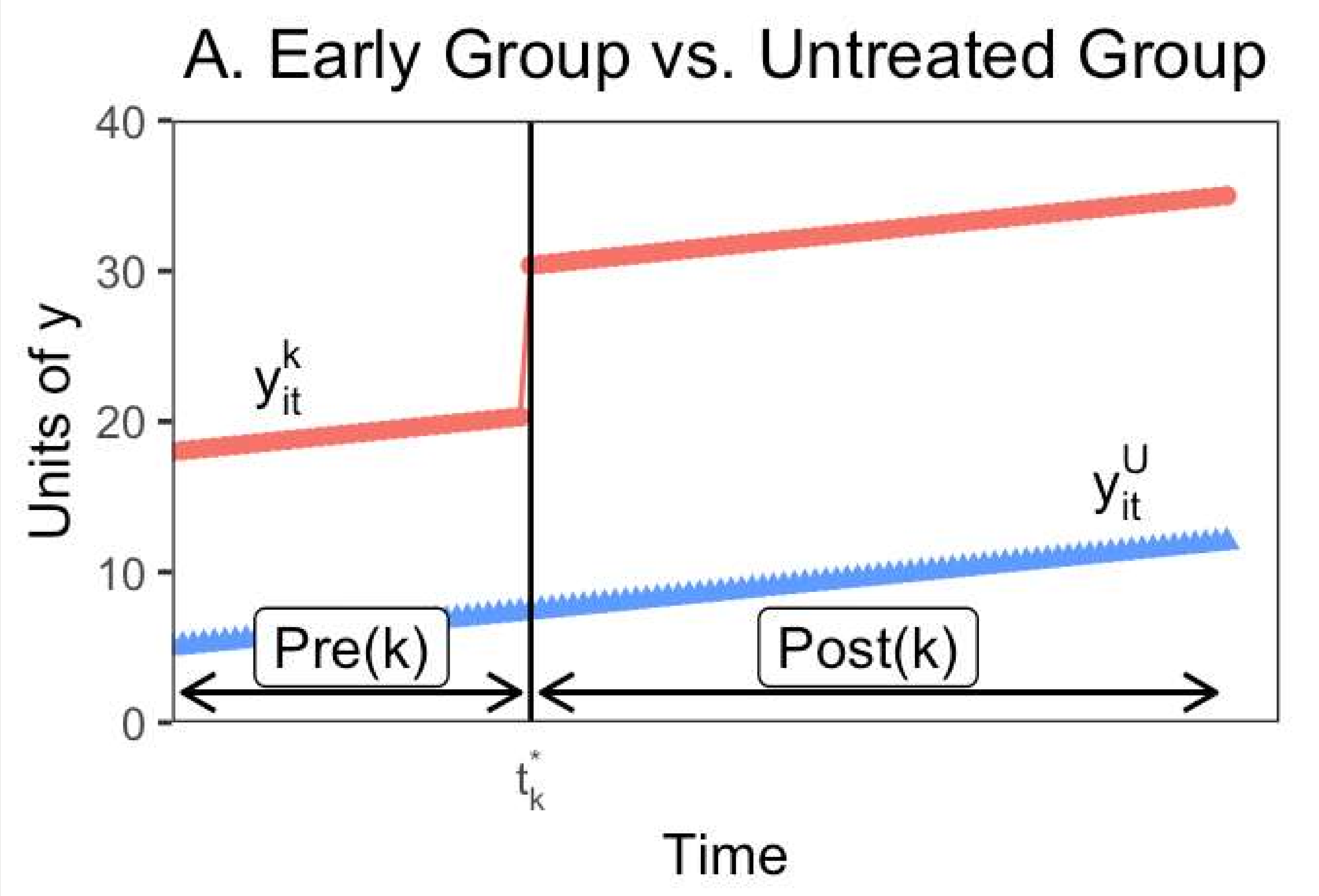
But what is β_{it}^{DD} actually telling us?

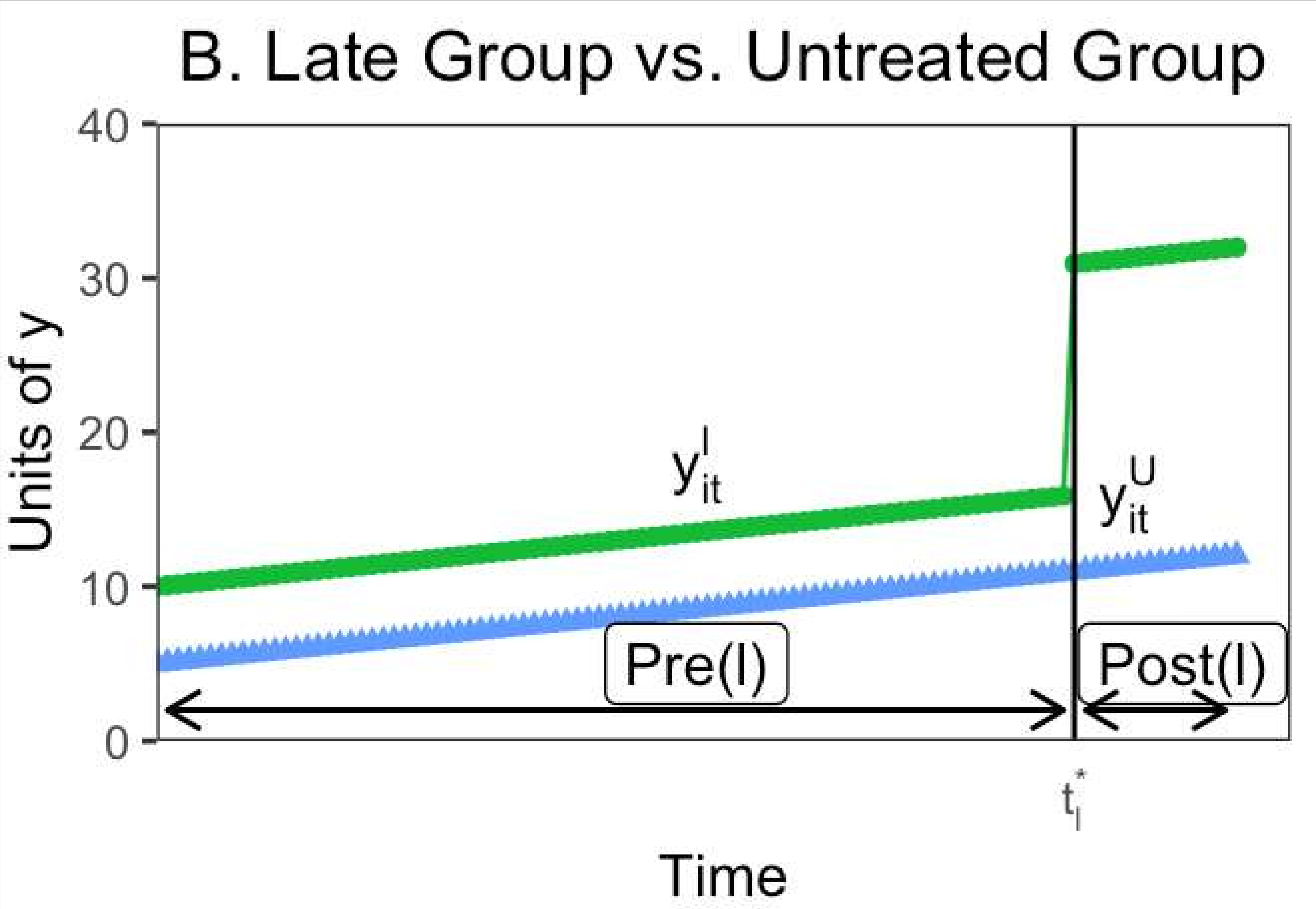
- For the canonical 2x2, we know exactly what we are estimating
- For *i* groups and *t* time periods, we are getting some average of multiple 2x2s
- But how does this work, exactly?

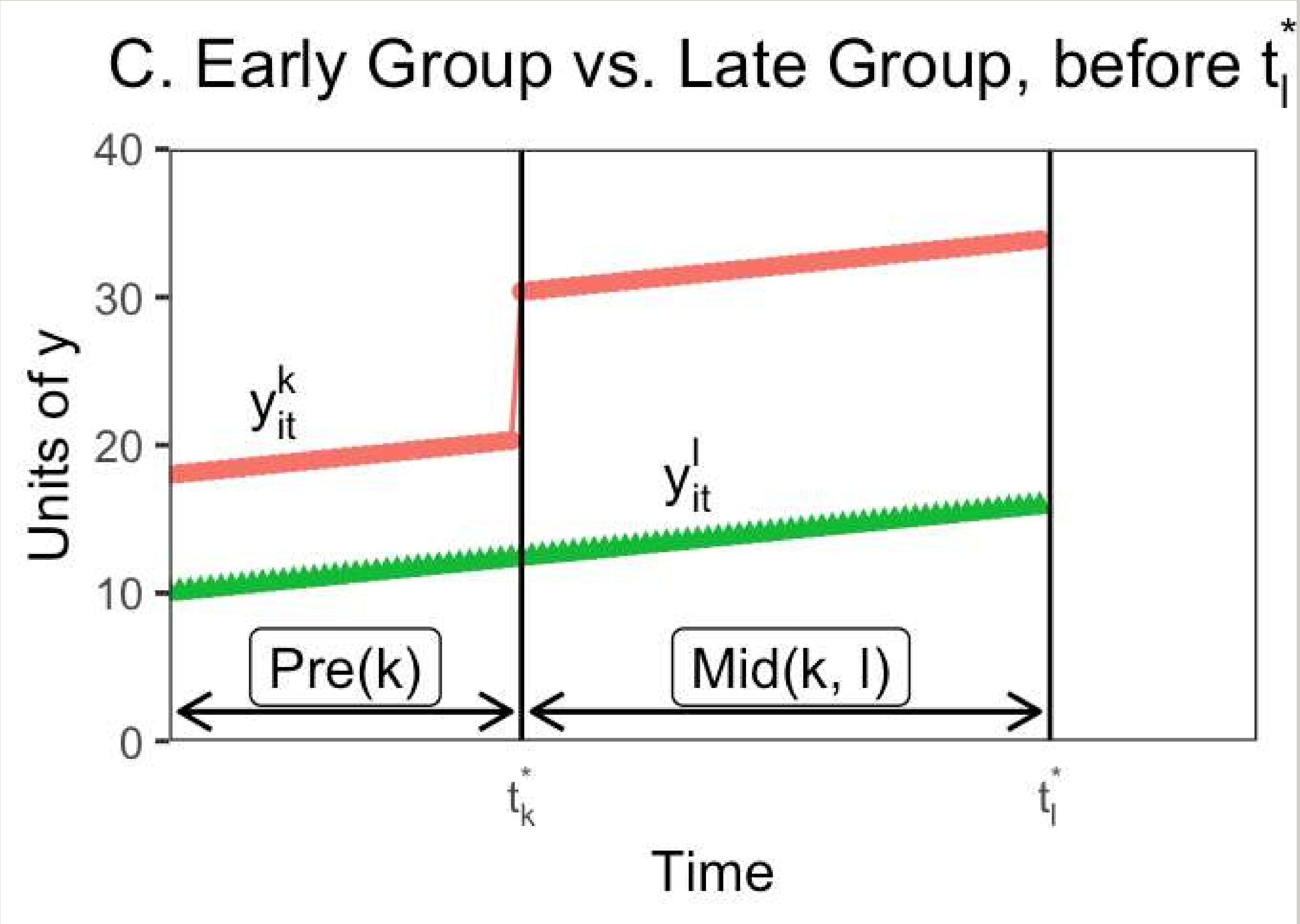
Let's take a single step from two time periods to three, where treatment can be adopted at either t = 2 or t = 3



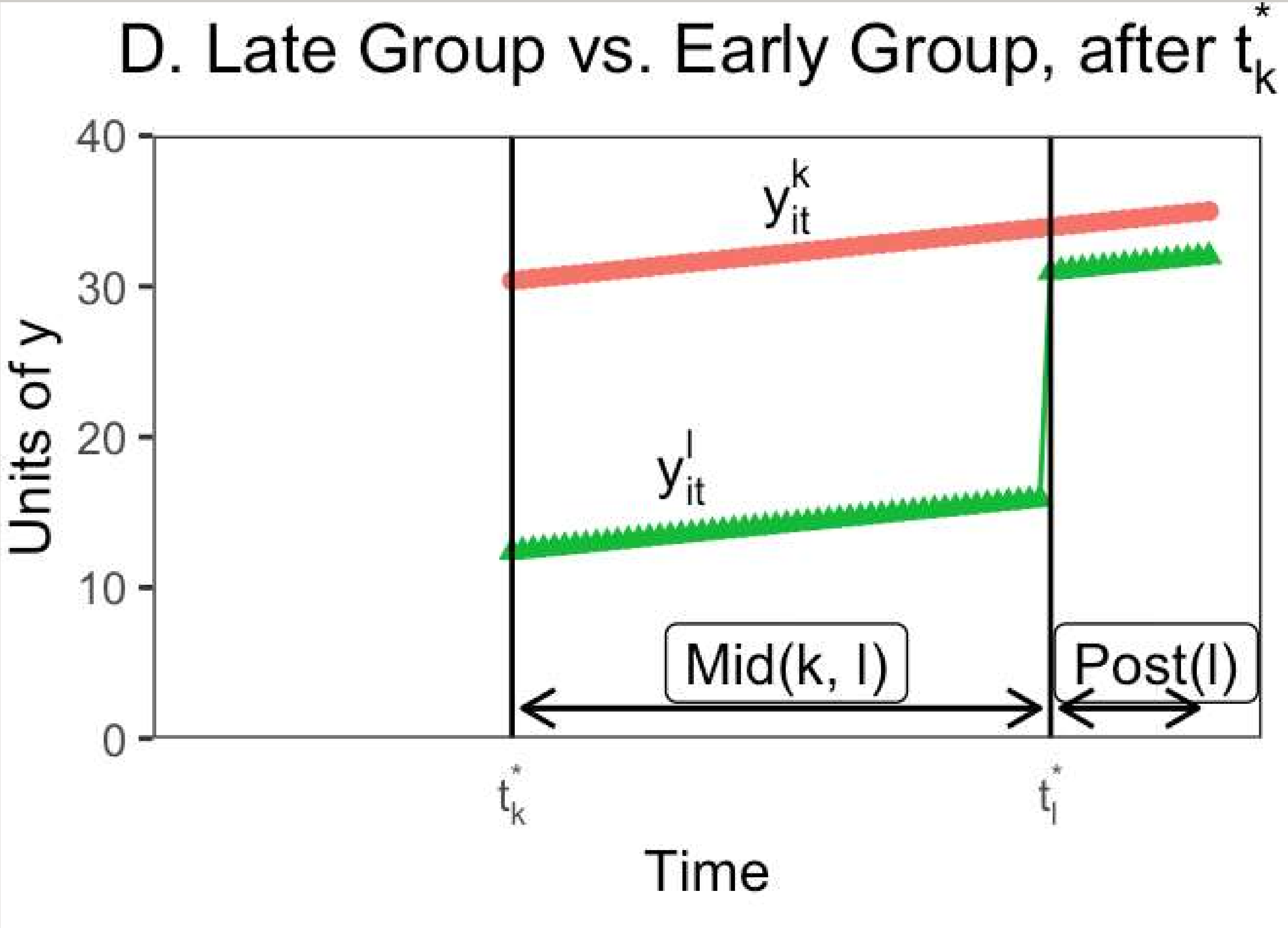
Any design with multiple treatment timings will have k^2 groups, where *k* is the number of timings.


A. Early Group vs. Untreated Group
Units of y
40
30
20
10
0
y_{it}^{k}
y_{it}^{U}
Pre(k)
Post(k)
t_k^*
Time


B. Late Group vs. Untreated Group
Units of y
40
30
20
10
0
y^{l}_{it}
y^{U}_{it}
Pre(l)
Post(l)
t^{*}_{l}
Time



C. Early Group vs. Late Group, before t_l^*
40
30
20
10
0
Units of y
y_{it}^k
y_{it}^l
Pre(k)
Mid(k, l)
t_k^*
t_l^*
Time



D. Late Group vs. Early Group, after t_k^*
40
30
20
10
0
Units of y
y_{it}^k
y_{it}^l
Mid(k, l)
Post(l)
t_k^*
t_l^*
Time

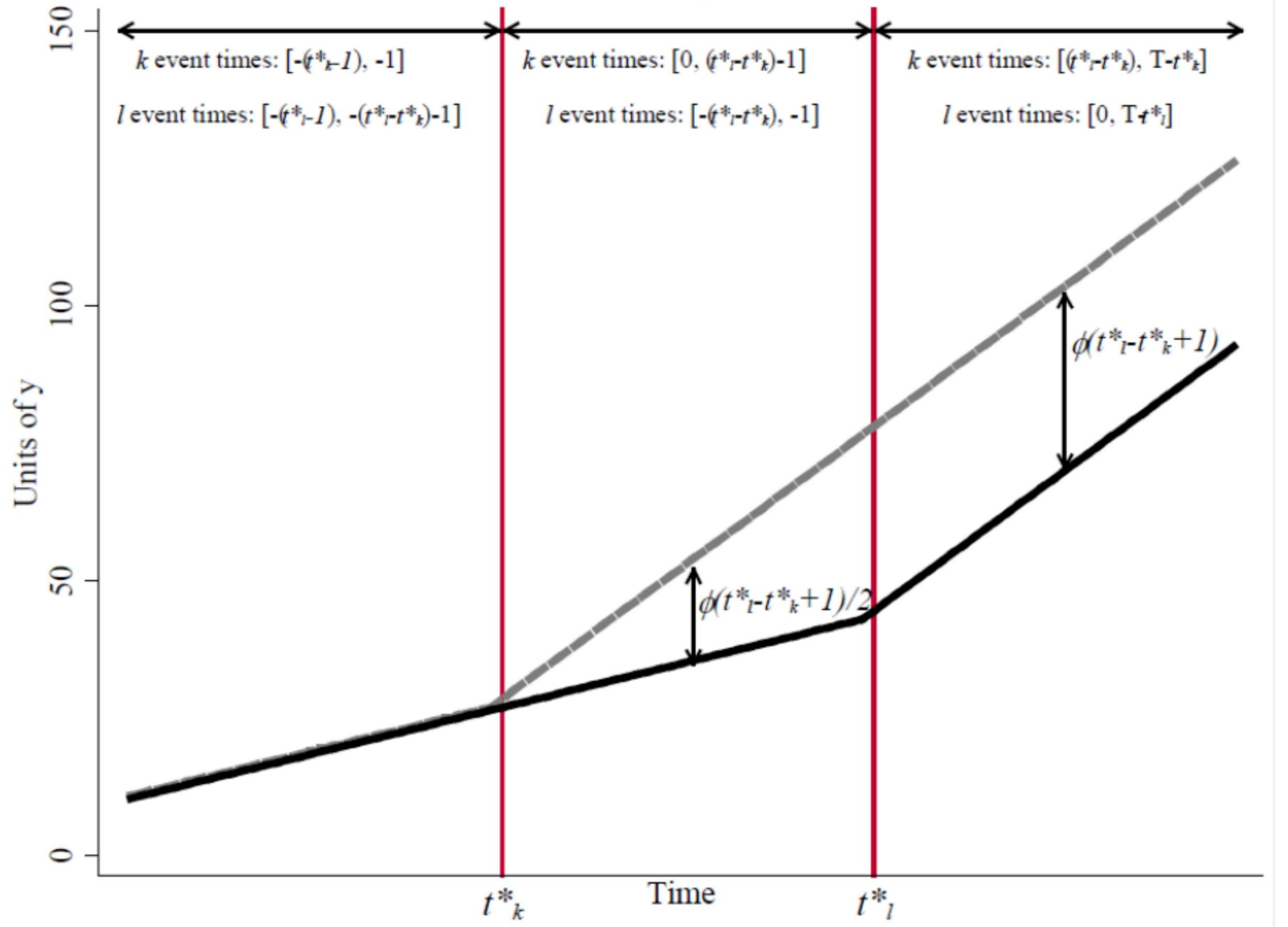
Where does this leave us?

- 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

But TWFE fails under 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

Figure 3. Difference-in-Differences Estimates with Variation in Timing Are Biased When Treatment Effects Vary Over Time



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

Adjustment: new estimators

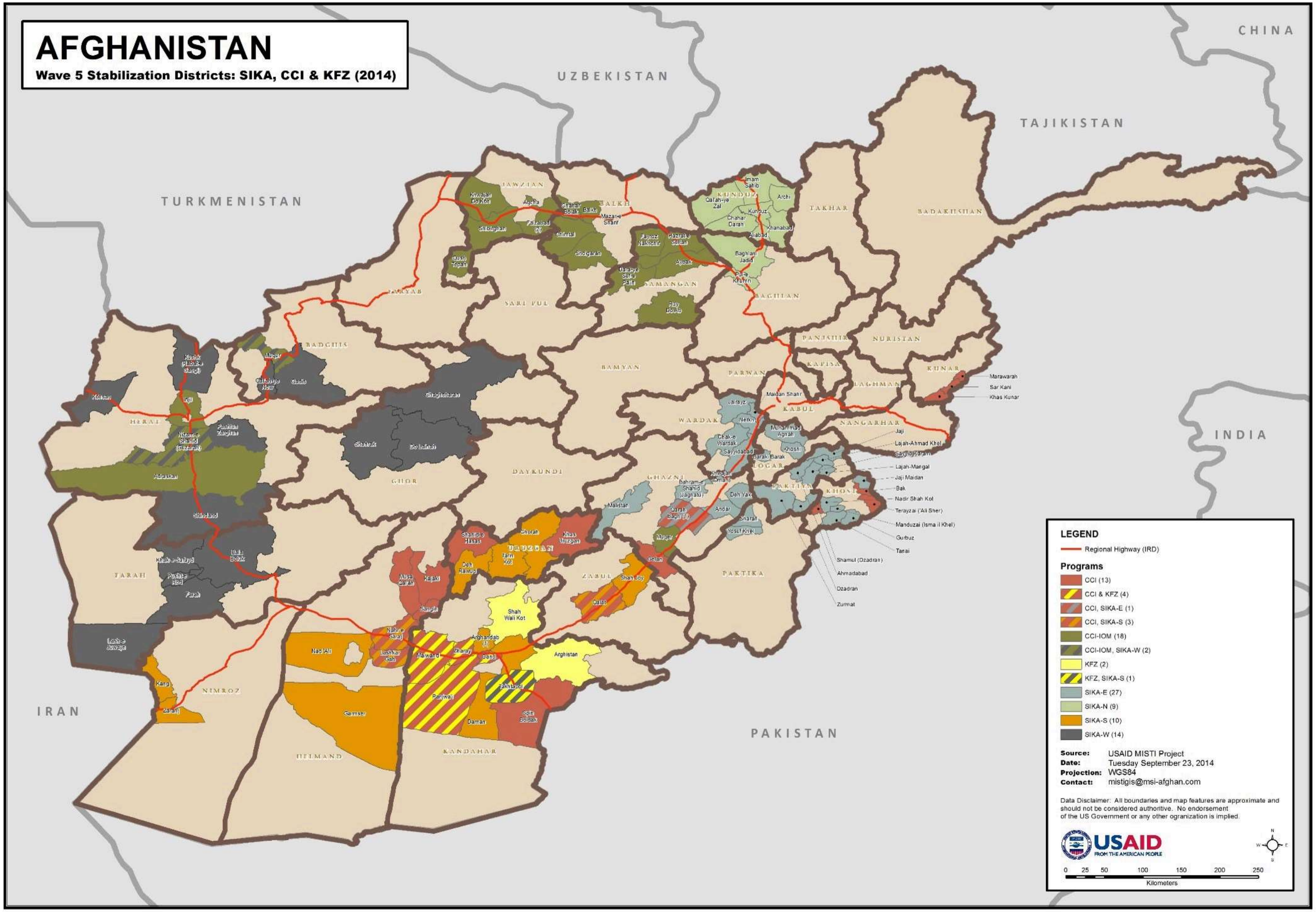
- Background
- Problem
- Solutions
- **Case study - MISTI**
- Final thoughts

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
TAJIKISTAN
CHINA
TURKMENISTAN
INDIA
IRAN
PAKISTAN
JAWZJAN
BALKH
KUNDUZ
TAKHAR
BADAKHSHAN
SAMANGAN
BAGHLAN
SARI PUL
FARYAB
BADGHIS
BAMYAN
PANJSHIR
NURISTAN
KAPISA
PARWAN
KUNAR
LAGHMAN
KABUL
WARDAK
NANGARHAR
HERAT
GHOR
DAYKUNDI
GHAZNI
LOGAR
PAKTIYA
KHOST
URUZGAN
ZABUL
PAKTIKA
FARAH
NIMROZ
HELMAND
KANDAHAR
Marawarah
Sar Kani
Khas Kunar
Jaji
Lajah-Ahmad Khel
Lajah-Mangal
Jaji Maidan
Bak
Nadir Shah Kot
Terayzai ('Ali Sher)
Manduzai (Isma il Khel)
Gurbuz
Tanai
Shamul (Dzadran)
Ahmadabad
Dzadran
Zurmat
Shah Wali Kot
Arghistan
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	Treated villages (cumulative)
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	0.031	0.043	0.003	-0.039	-0.002

MISTI Two Way Fixed Effects (TWFE)

$$y_{it} = \alpha_i + \alpha_t + \beta_{it}^{DD} + \epsilon_{it}$$

$$y_{it} = village_i + wave_t + treated_{it}^{DD} + \epsilon_{it}$$

term	estimate	std.error	statistic	p.value
treat_event	-0.0389	0.0947	-0.411	0.681

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

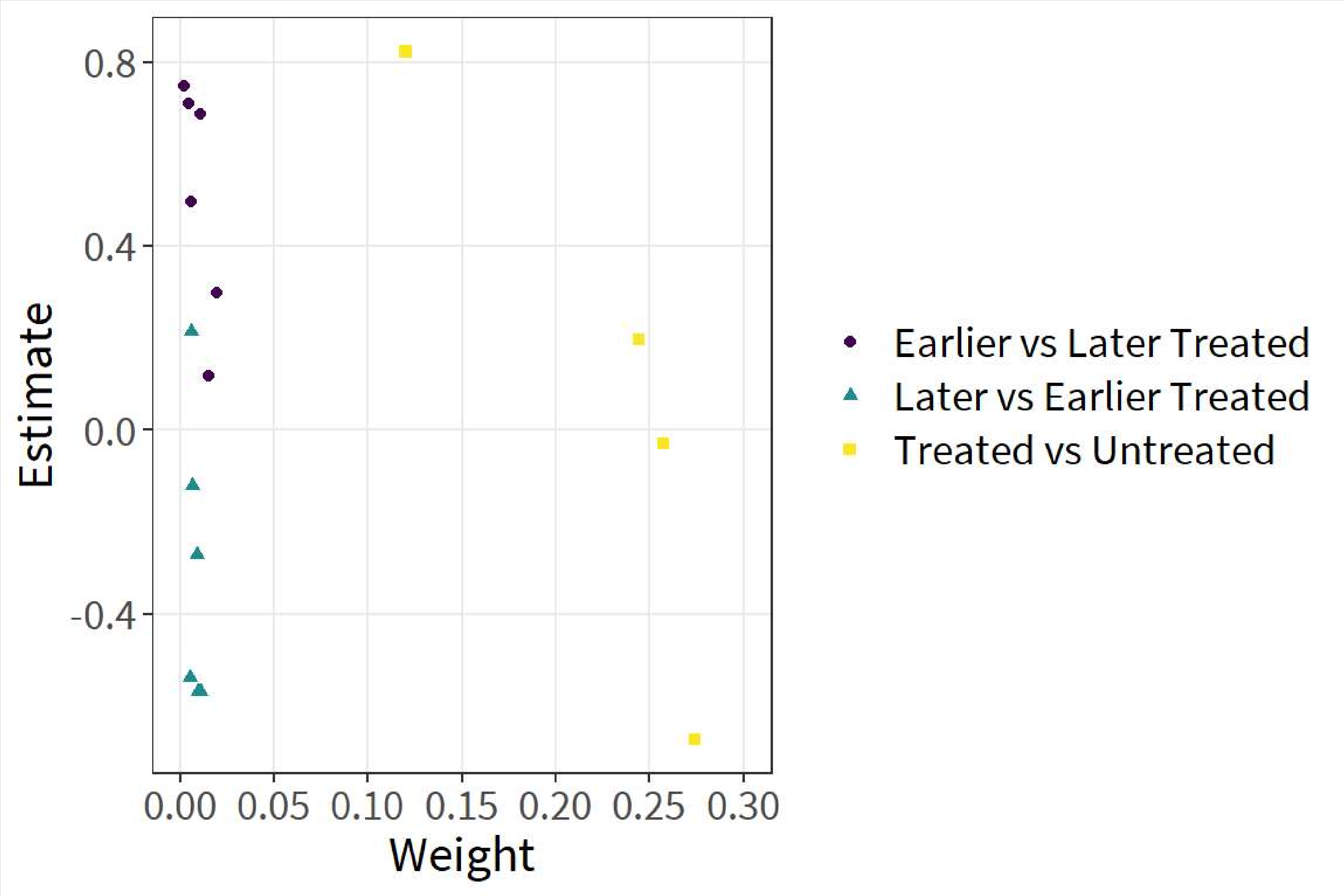
```
                          type weight avg_est
1 Earlier vs Later Treated 0.0578  0.3911
2 Later vs Earlier Treated 0.0466 -0.3465
3     Treated vs Untreated 0.8956 -0.0506
```

MISTI bacondecomp 2x2 cells

treated	untreated	estimate	weight	type
2	3	0.7473	0.00211	Earlier vs Later Treated
2	4	0.7093	0.00444	Earlier vs Later Treated
2	5	0.6868	0.01066	Earlier vs Later Treated
2	99,999	0.8232	0.11998	Treated vs Untreated
3	2	-0.1216	0.00633	Later vs Earlier Treated
3	4	0.4973	0.00603	Earlier vs Later Treated
3	5	0.2976	0.01930	Earlier vs Later Treated
3	99,999	0.1964	0.24425	Treated vs Untreated
4	2	-0.2707	0.00889	Later vs Earlier Treated
4	3	0.2135	0.00603	Later vs Earlier Treated
4	5	0.1182	0.01524	Earlier vs Later Treated
4	99,999	-0.0291	0.25710	Treated vs Untreated
5	2	-0.5680	0.01066	Later vs Earlier Treated
5	3	-0.5686	0.00965	Later vs Earlier Treated
5	4	-0.5375	0.00508	Later vs Earlier Treated
5	99,999	-0.6729	0.27424	Treated vs Untreated

```
[1] -0.0389
```

Plot of 2x2 cells

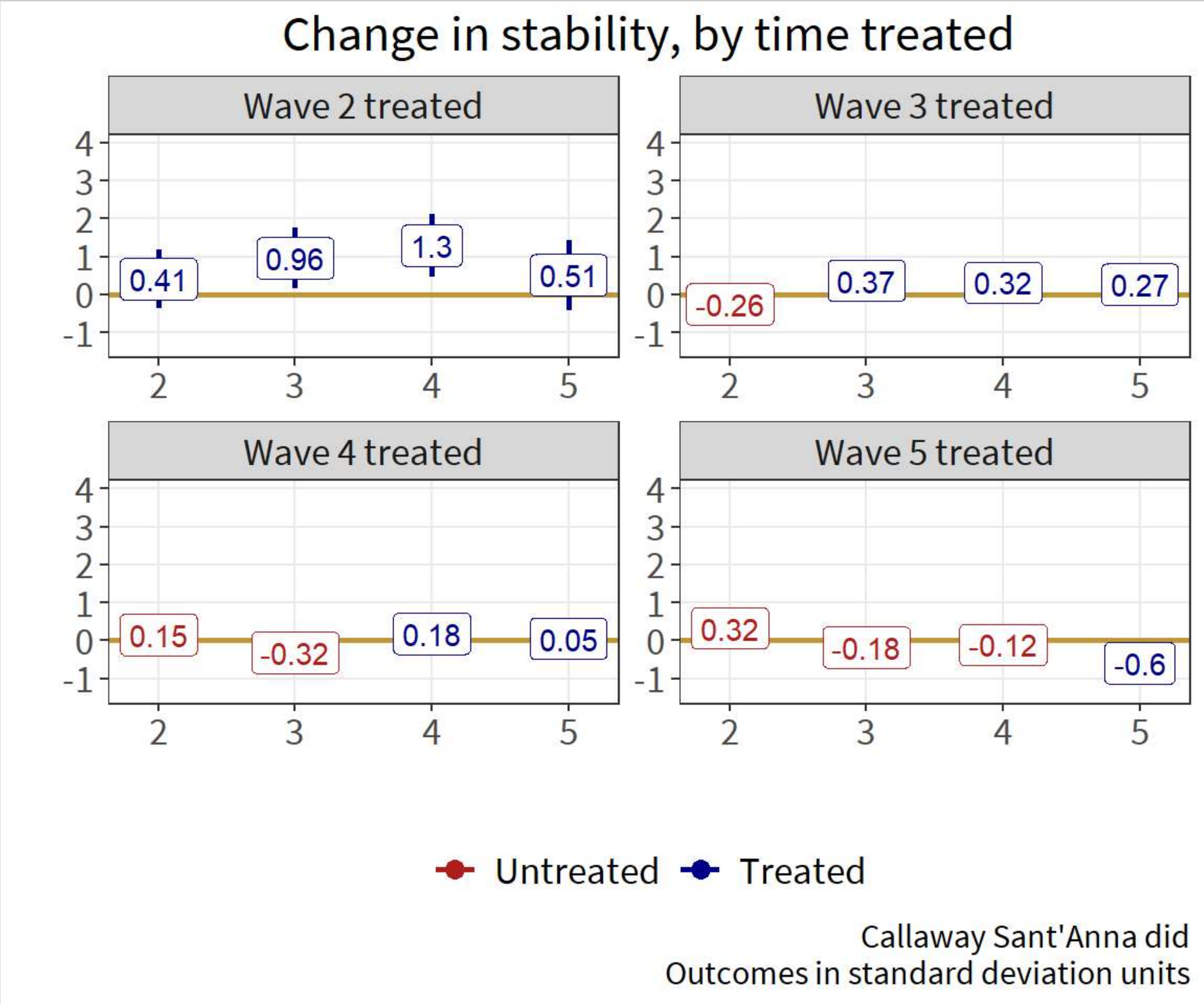


Callaway and Sant'Anna (2020)

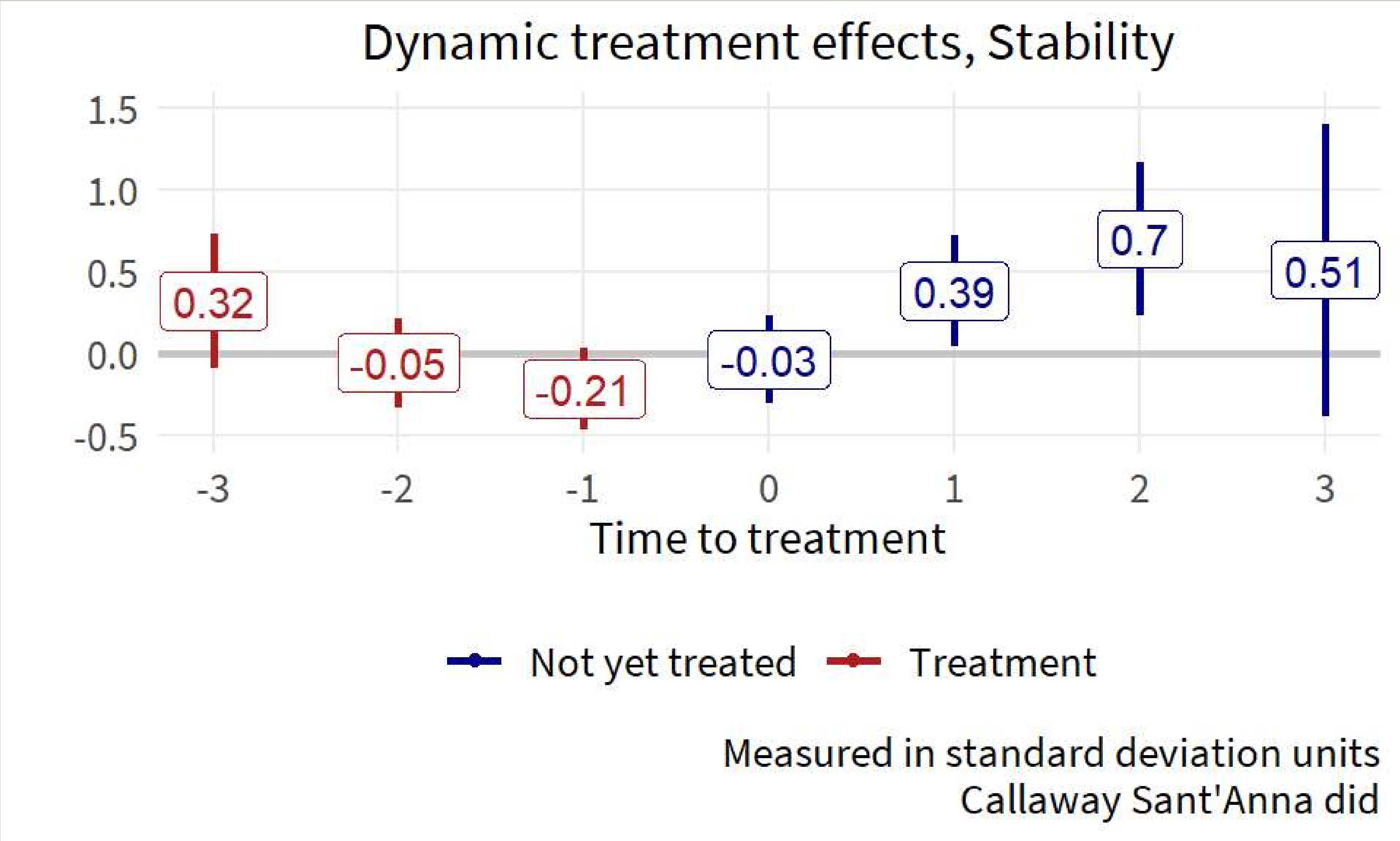
This estimation gives you multiple outcomes

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

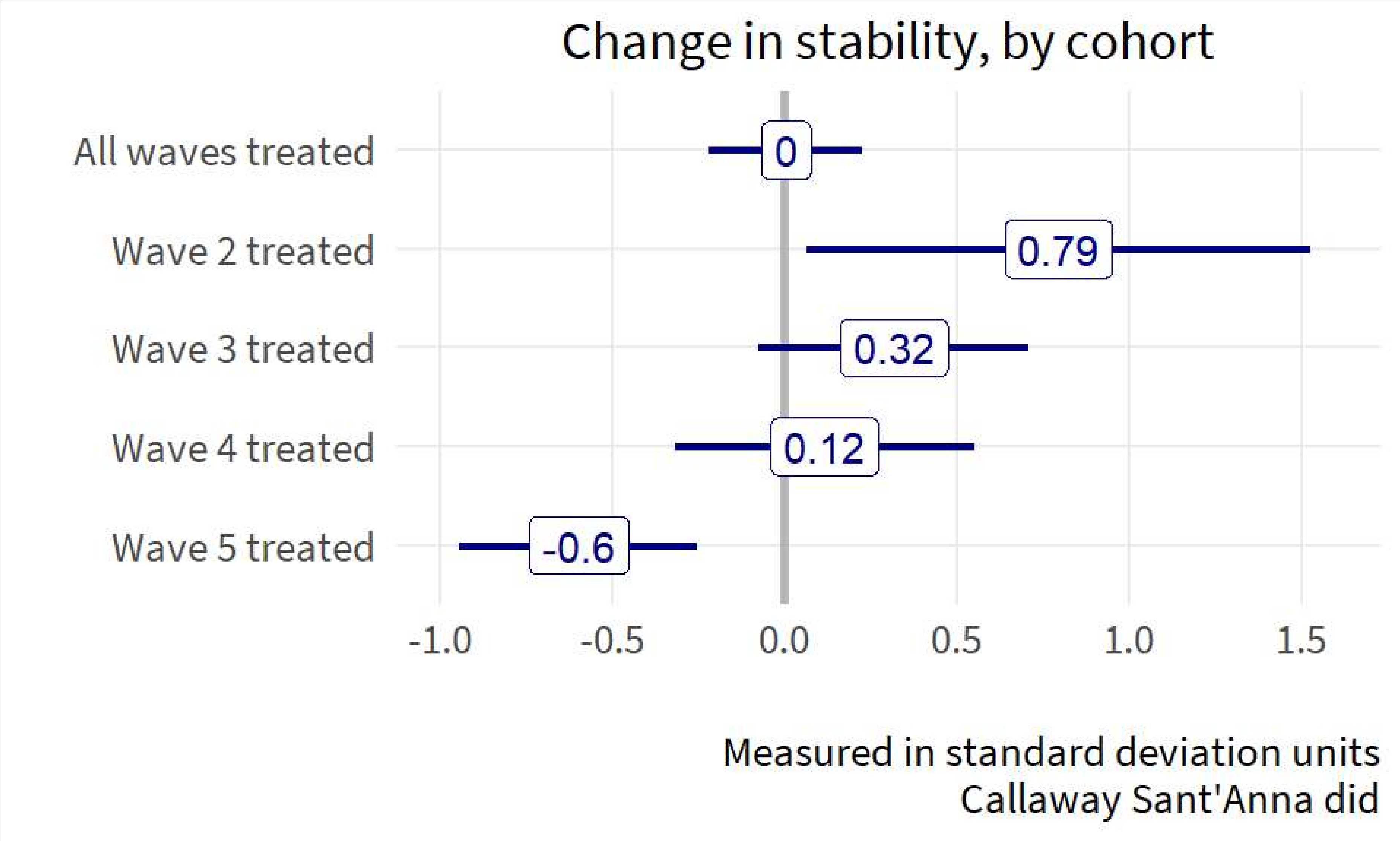
Treatment effects for each wave



Dynamic treatment effects



Overall effects by cohort



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

What have we learned?

- 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!