A trial emulation approach for policy evaluations with group-level longitudinal data

Eli Ben-Michael
Harvard University

Joint work with Avi Feller and Elizabeth Stuart

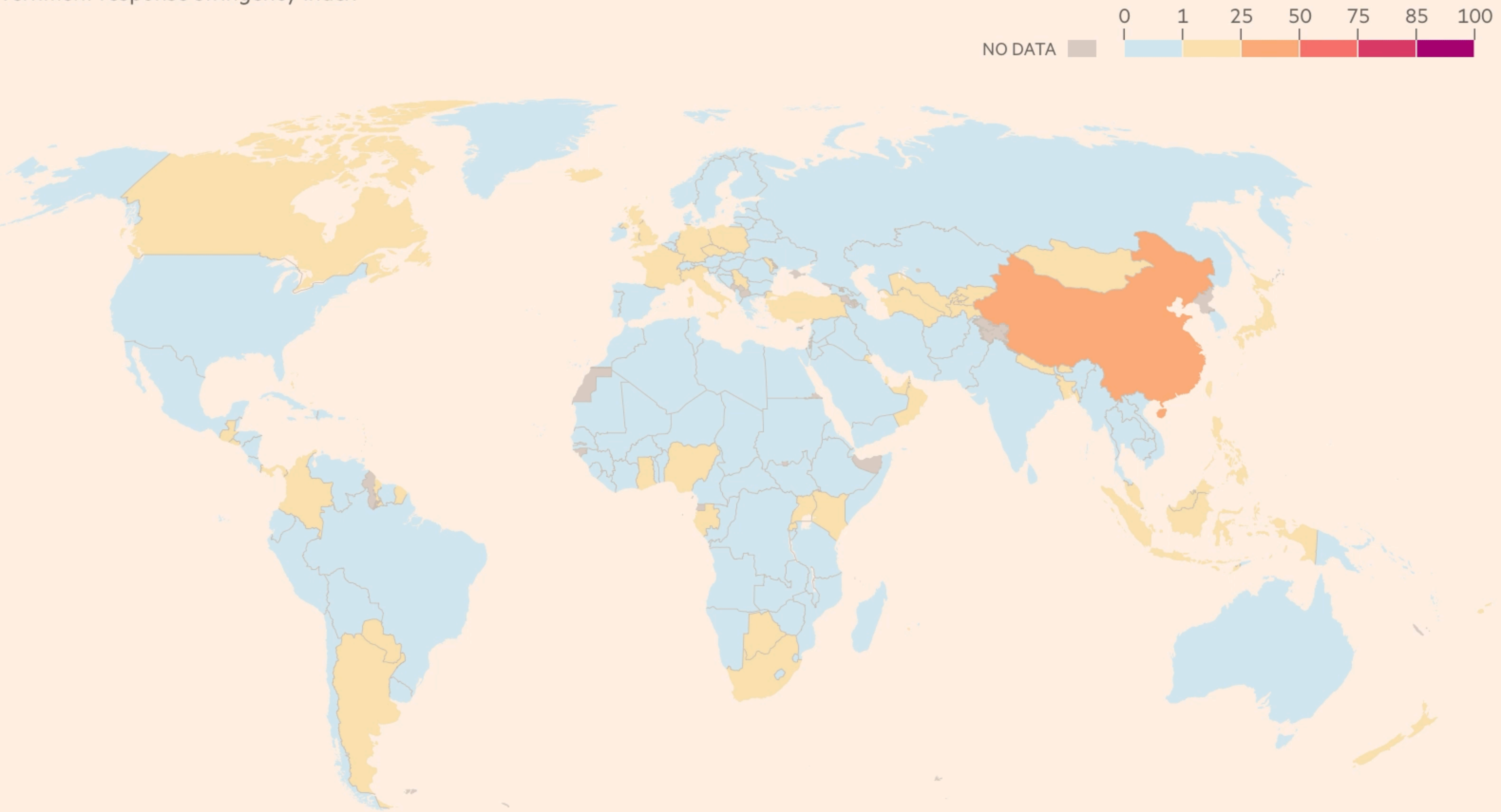
2020 was an extraordinary year (for policy evaluation)

- Unprecedented policy measures
- Wide variation in types of NPIs
- Rapidly changing policy environment

Jan 24 2020

Lockdowns around the world

Oxford Covid-19 government response stringency index



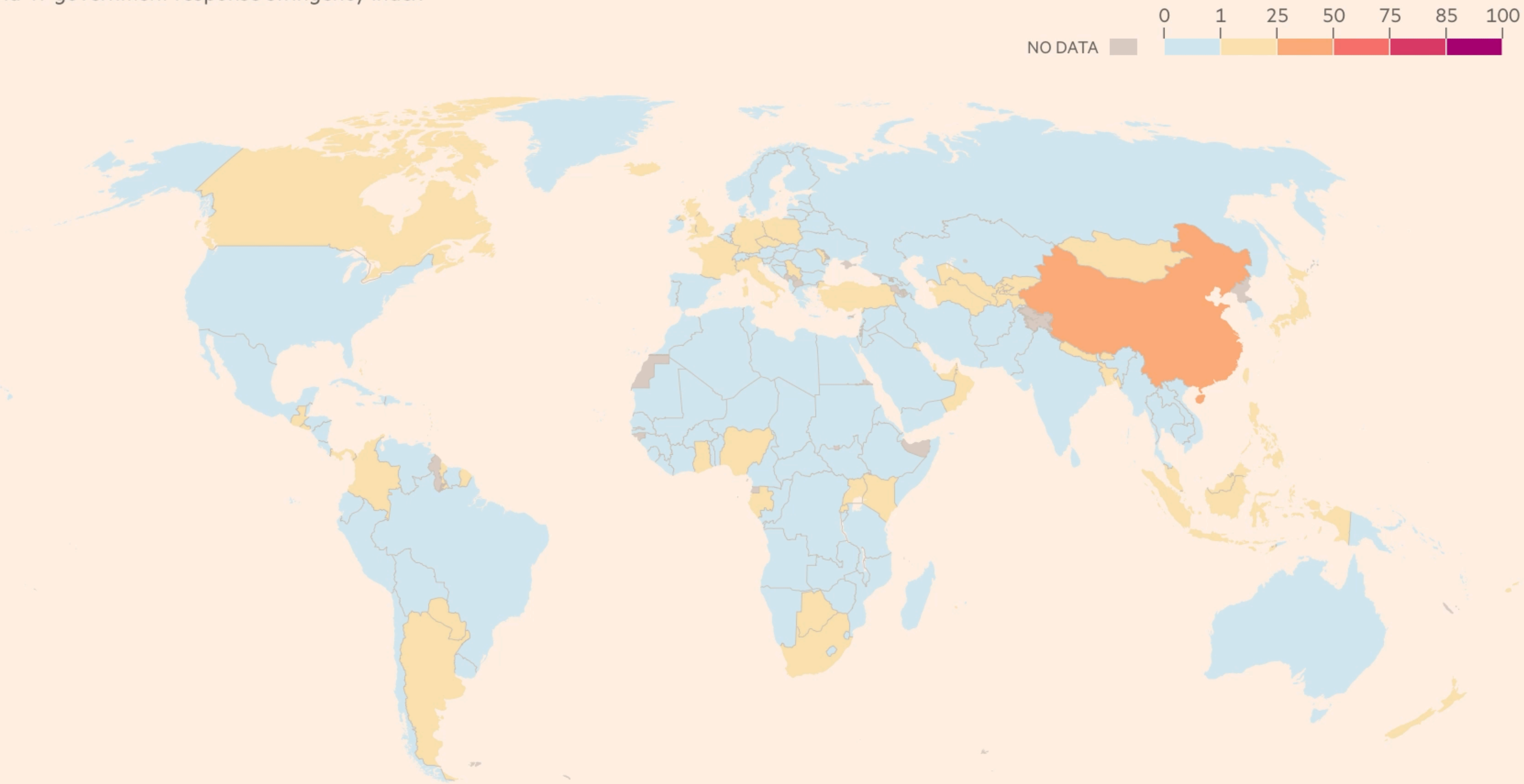
Graphic: Alan Smith and David Blood

Source: Blavatnik School of Government, University of Oxford. Data as of April 22. Data for the most recent seven days may not yet reflect government response changes implemented during that period

Jan 24 2020

Lockdowns around the world

Oxford Covid-19 government response stringency index



Graphic: Alan Smith and David Blood
Source: Blavatnik School of Government, University of Oxford. Data as of April 22. Data for the most recent seven days may not yet reflect government response changes implemented during that period

Policy evaluation is hard!

…especially during Covid-19

What's so difficult?

(an incomplete list)

- Policies are not randomized
- Policies are adopted at different times
- Multiple policies are bundled together
- Policies do not determine individual behavior
- Policies in one location might affect another

But it's important to evaluate policy impacts!

Design-based thinking can help

Design-based thinking can help

Target Trial Emulation

Design an observational study like a randomized one

[Danaei et al 2018; Dickerman et al 2019]

Design-based thinking can help

Target Trial Emulation

Design an observational study like a randomized one

[Danaei et al 2018; Dickerman et al 2019]

Panel Data Methods

Beyond two-way fixed effects

[Goodman-Bacon 2018; Abraham & Sun 2021; Callaway & Sant'Anna 2021]

Design-based thinking can help

Design-based thinking can help

Combines insights from Epidemiology and Econometrics

Design-based thinking can help

Target Trial Emulation

Design an observational study like a randomized one

[Danaei et al 2018; Dickerman et al 2019]

Panel Data Methods

Beyond two-way fixed effects

[Goodman-Bacon 2018; Abraham & Sun 2021; Callaway & Sant'Anna 2021]

Policy Trial Emulation

Combines insights from Epidemiology and Econometrics

A stylized analysis:

- Evaluate the impact of stay at home orders in the US

Outline

1. The elements of policy trial emulation
2. Single and nested target trials

Policy Trial Emulation

The elements of policy trial emulation

The elements of policy trial emulation

Require 4 definitions:

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts
3. Outcomes

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts
3. Outcomes
4. Time zero

The elements of policy trial emulation

Require 4 definitions:

1. **Units and exposures**
2. Causal contrasts
3. Outcomes
4. Time zero

The elements of policy trial emulation

Require 4 definitions:

1. **Units and exposures**
2. Causal contrasts
3. Outcomes
4. Time zero

- What is a stay-at-home policy?
  - Variation across states
  - We'll package these all together → less interpretable

The elements of policy trial emulation

Require 4 definitions:

1. **Units and exposures**
2. Causal contrasts
3. Outcomes
4. Time zero

- What is a stay-at-home policy?
  - Variation across states
  - We'll package these all together → less interpretable
- Are we interested in policies or behavior?
  - Individual mobility reduced **before** policy changes
  - Limited compliance and enforcement

[Goolsbee & Syverson 2020]

The elements of policy trial emulation

Require 4 definitions:

1. **Units and exposures**
2. Causal contrasts
3. Outcomes
4. Time zero

- What is a stay-at-home policy?
  - Variation across states
  - We'll package these all together → less interpretable
- Are we interested in policies or behavior?
  - Individual mobility reduced **before** policy changes
  - Limited compliance and enforcement

[Goolsbee & Syverson 2020]

- Are there spillovers?
  - Probably! But this is difficult to account for

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. **Causal contrasts**
3. Outcomes
4. Time zero

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. **Causal contrasts**
3. Outcomes
4. Time zero

- Intent-to-treat (ITT) analysis
  - Measure the effect of the policy as implemented

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. **Causal contrasts**
3. Outcomes
4. Time zero

- Intent-to-treat (ITT) analysis
  - Measure the effect of the policy as implemented
- Potential outcomes framework
  - W_{it} State i has a stay-at-home order at time t
  - $Y_{it}(1), Y_{it}(0)$ Outcome if order is/isn't enacted
  - Average of instantaneous effects $Y_{it}(1) - Y_{it}(0)$ for states that enacted a stay at home policy

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. **Causal contrasts**
3. Outcomes
4. Time zero

- Intent-to-treat (ITT) analysis
  - Measure the effect of the policy as implemented
- Potential outcomes framework
  - W_{it} State i has a stay-at-home order at time t
  - $Y_{it}(1), Y_{it}(0)$ Outcome if order is/isn't enacted
  - Average of instantaneous effects $Y_{it}(1) - Y_{it}(0)$ for states that enacted a stay at home policy
- Only focus on starting stay-at-home orders
  - Effect of "turning off" policies adds complexity

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts
3. **Outcomes**
4. Time zero

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts
3. **Outcomes**
4. Time zero

- Cumulative vs instantaneous outcomes
  - Total number of Covid-19 cases
  - Ratio of current current case count to previous day

The elements of policy trial emulation

Require 4 definitions:

1. Units and exposures
2. Causal contrasts
3. **Outcomes**
4. Time zero

- Cumulative vs instantaneous outcomes
  - Total number of Covid-19 cases
  - Ratio of current current case count to previous day
- Transforming the outcomes
  - Logarithm transformation due to exponential growth

The elements of policy trial emulation

Require 4 definitions:

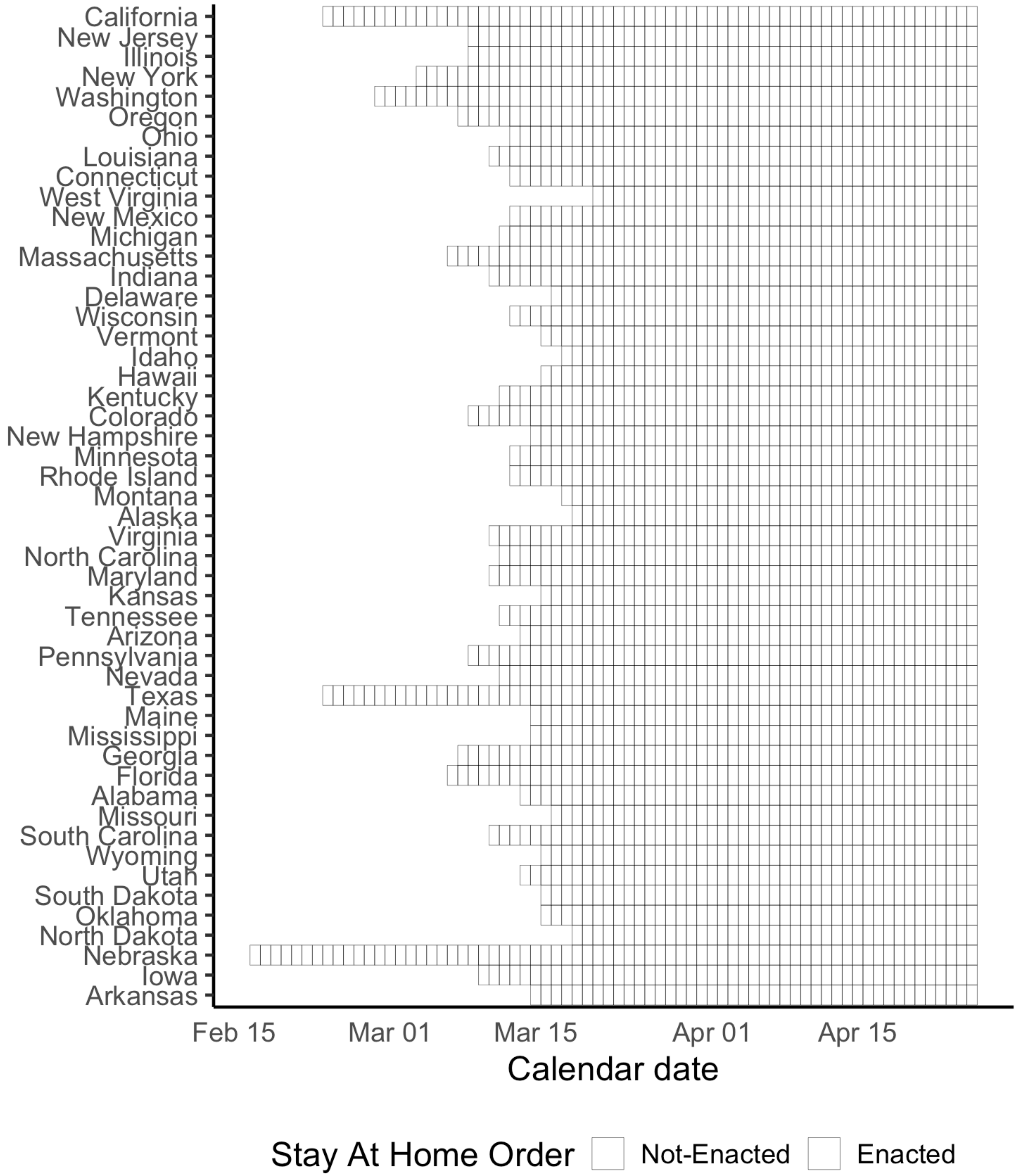
1. Units and exposures
2. Causal contrasts
3. **Outcomes**
4. Time zero

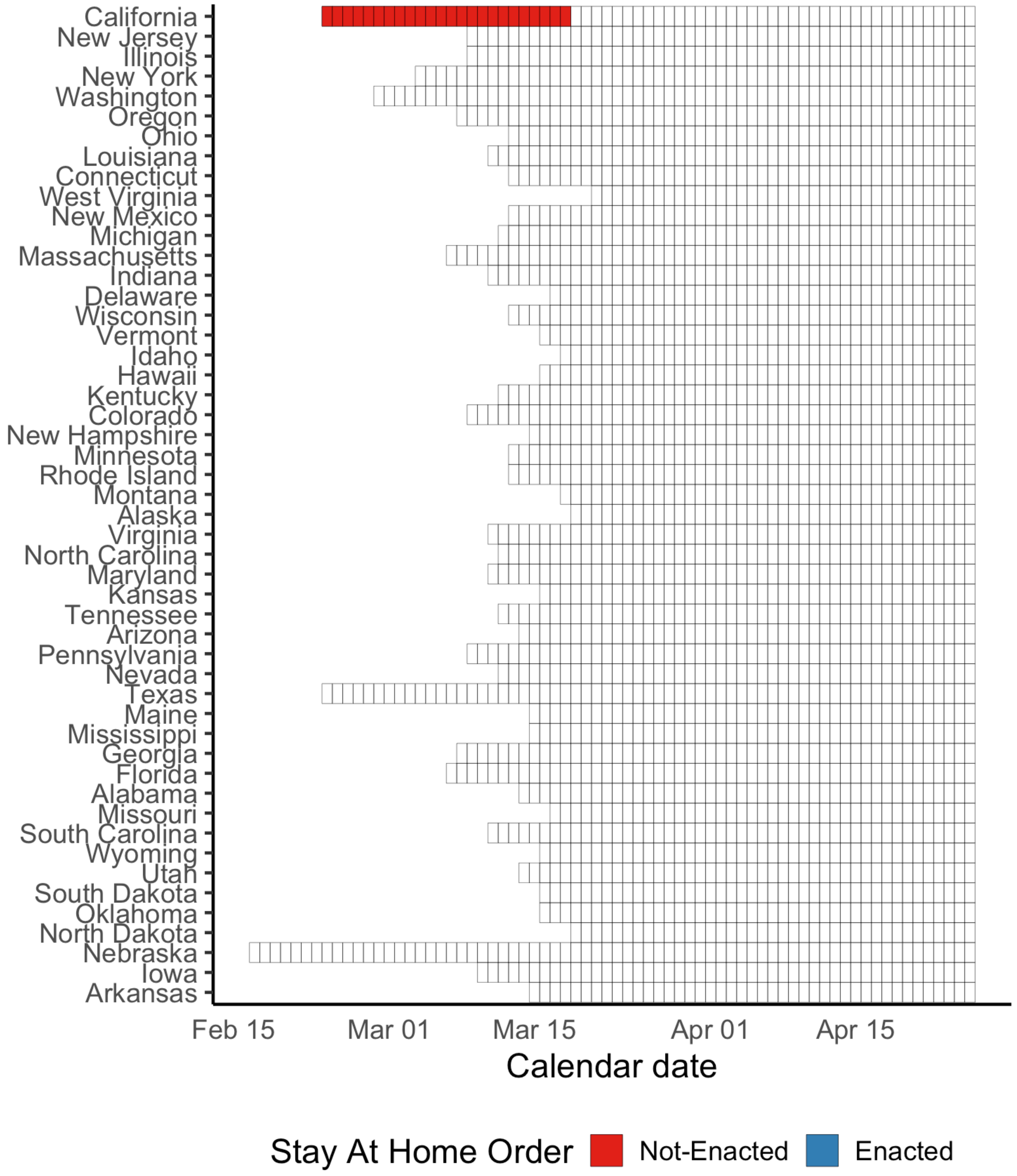
- Cumulative vs instantaneous outcomes
  - Total number of Covid-19 cases
  - Ratio of current current case count to previous day
- Transforming the outcomes
  - Logarithm transformation due to exponential growth
- Data quality is a concern
  - Differential changes in testing regimes over time?

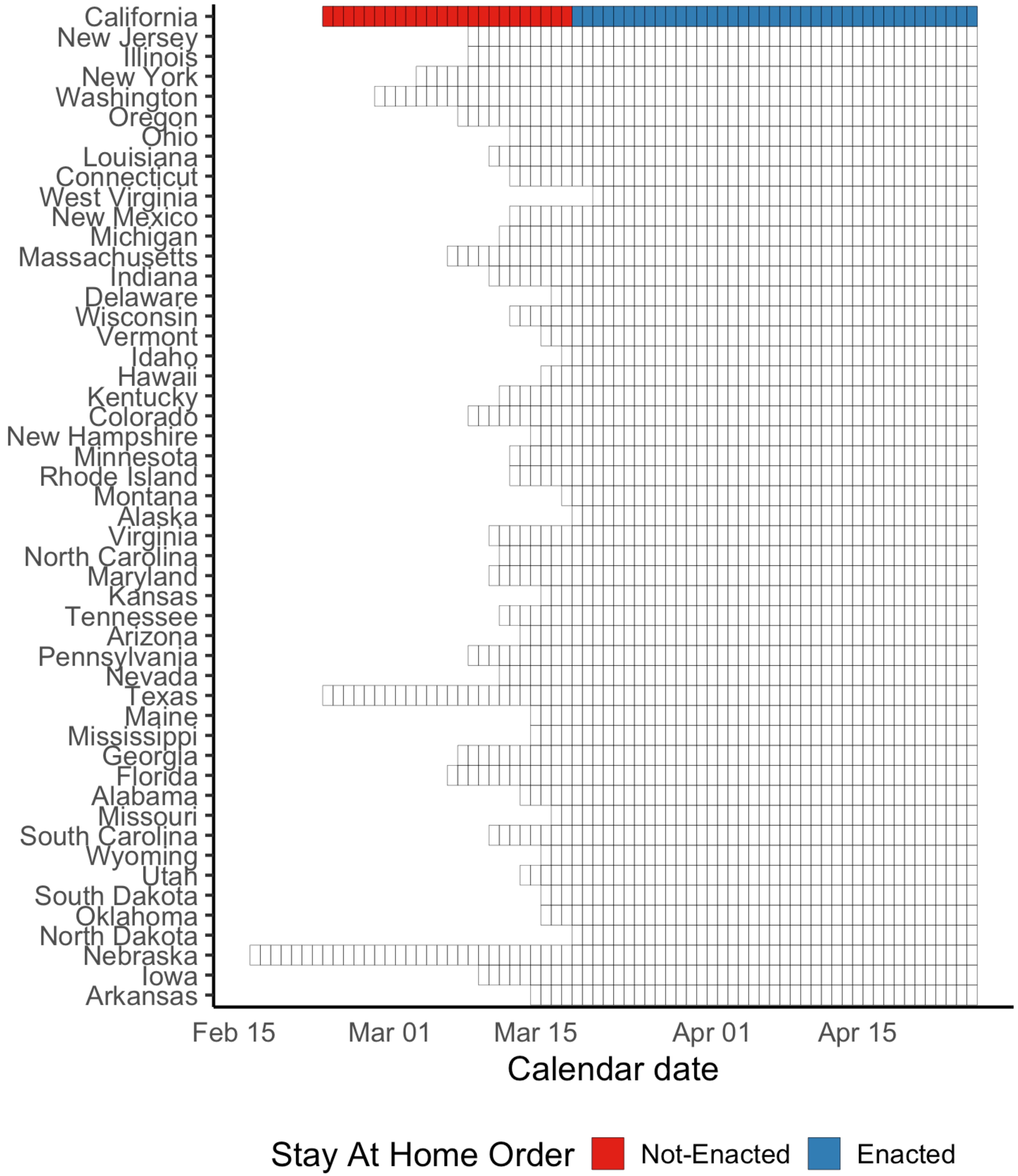
The elements of policy trial emulation

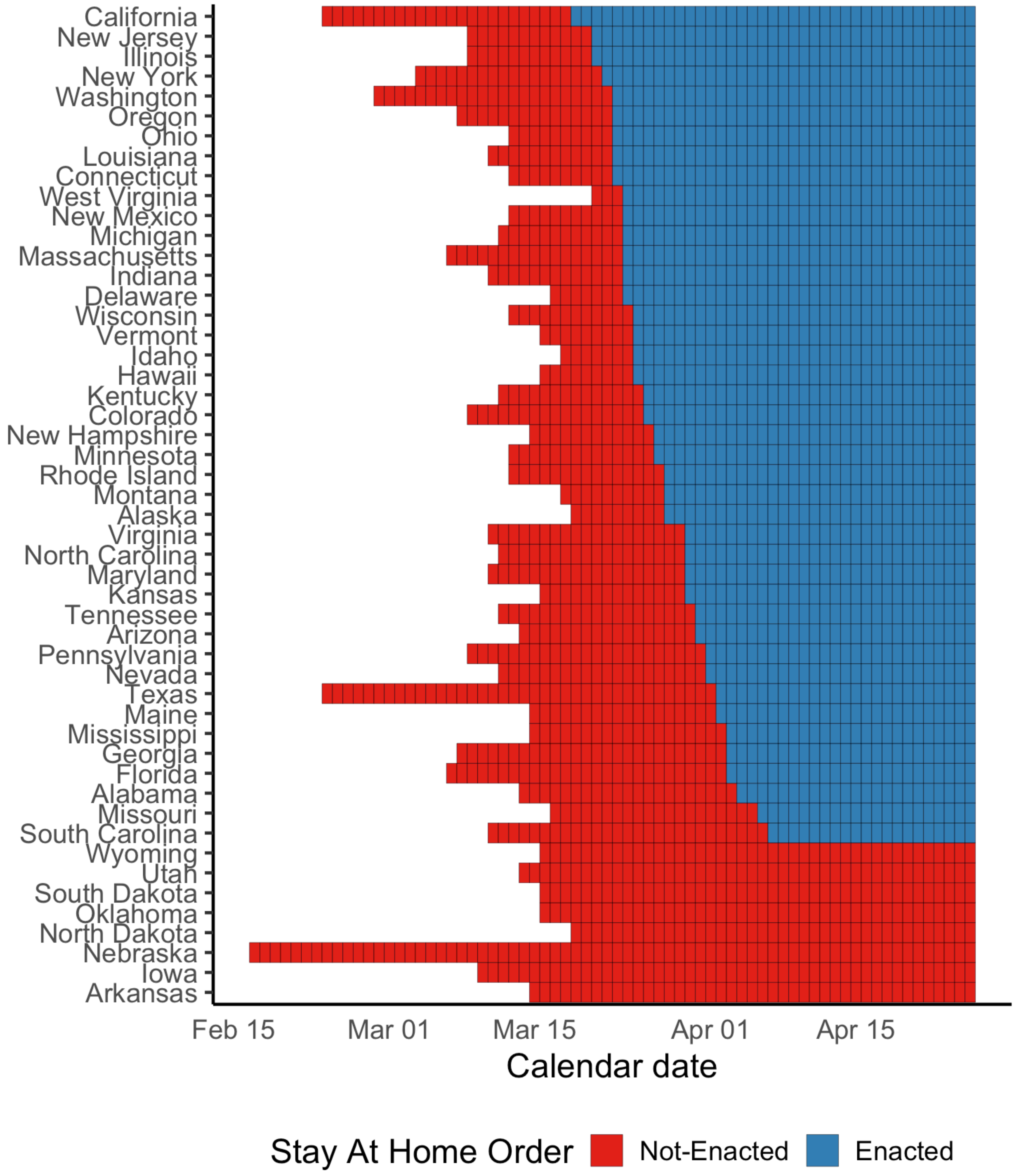
Require 4 definitions:

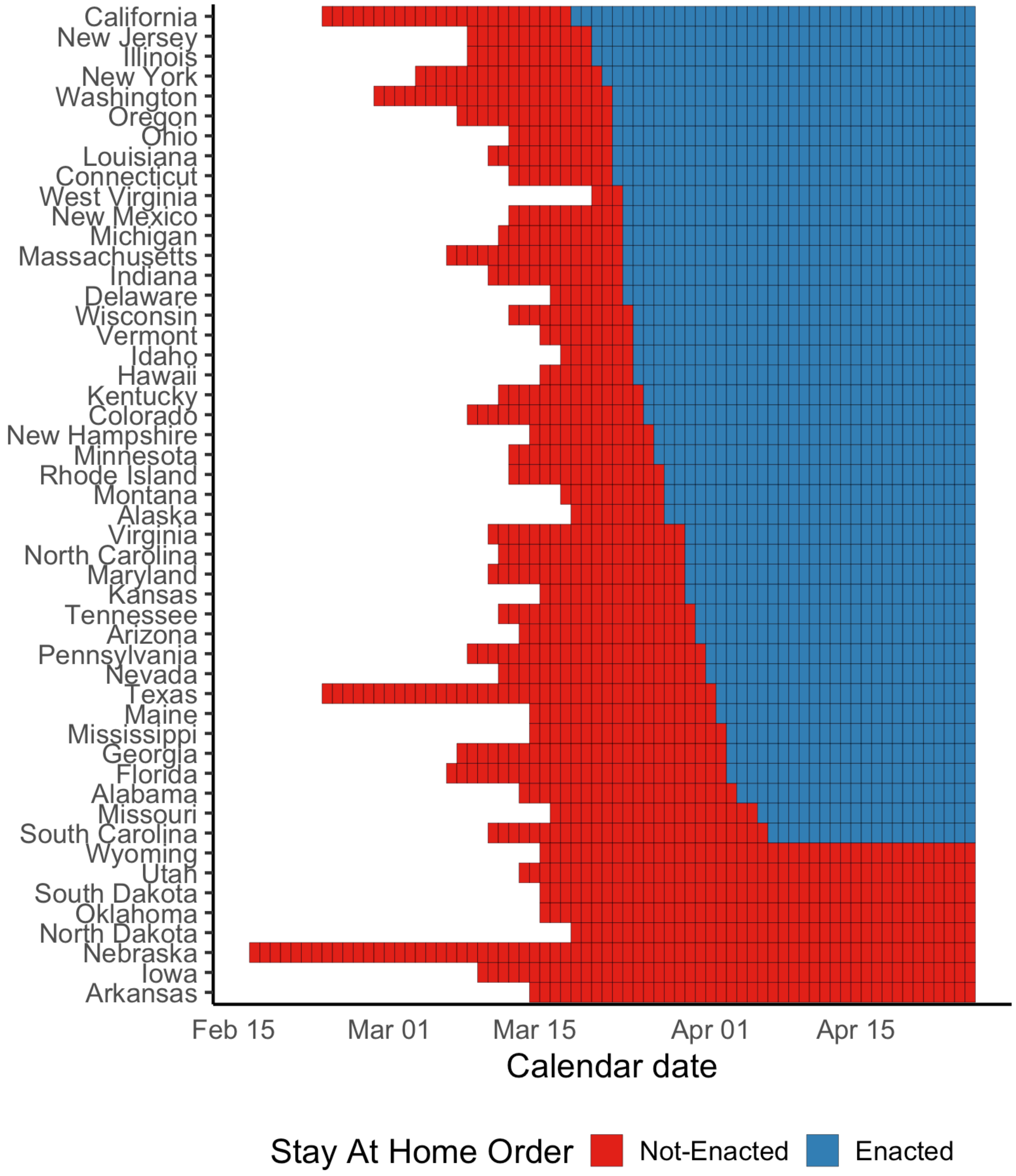
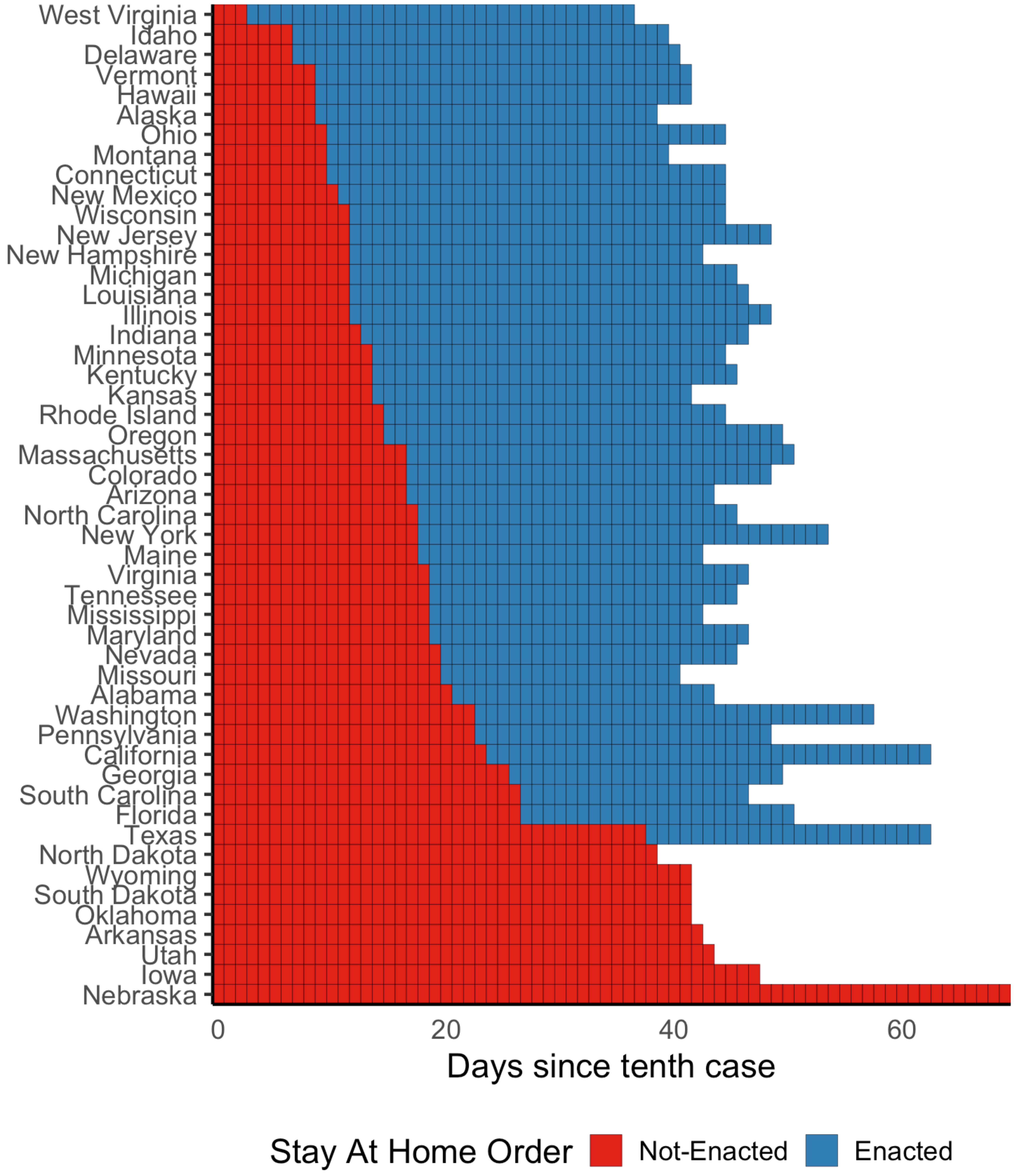
1. Units and exposures
2. Causal contrasts
3. Outcomes
4. **Time zero**


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted

West Virginia
Idaho
Delaware
Vermont
Hawaii
Alaska
Ohio
Montana
Connecticut
New Mexico
Wisconsin
New Jersey
New Hampshire
Michigan
Louisiana
Illinois
Indiana
Minnesota
Kentucky
Kansas
Rhode Island
Oregon
Massachusetts
Colorado
Arizona
North Carolina
New York
Maine
Virginia
Tennessee
Mississippi
Maryland
Nevada
Missouri
Alabama
Washington
Pennsylvania
California
Georgia
South Carolina
Florida
Texas
North Dakota
Wyoming
South Dakota
Oklahoma
Arkansas
Utah
Iowa
Nebraska
0
20
40
60
Days since tenth case
Stay At Home Order
Not-Enacted
Enacted


California
New Jersey
Illinois
New York
Washington
Oregon
Ohio
Louisiana
Connecticut
West Virginia
New Mexico
Michigan
Massachusetts
Indiana
Delaware
Wisconsin
Vermont
Idaho
Hawaii
Kentucky
Colorado
New Hampshire
Minnesota
Rhode Island
Montana
Alaska
Virginia
North Carolina
Maryland
Kansas
Tennessee
Arizona
Pennsylvania
Nevada
Texas
Maine
Mississippi
Georgia
Florida
Alabama
Missouri
South Carolina
Wyoming
Utah
South Dakota
Oklahoma
North Dakota
Nebraska
Iowa
Arkansas
Feb 15
Mar 01
Mar 15
Apr 01
Apr 15
Calendar date
Stay At Home Order
Not-Enacted
Enacted
West Virginia
Idaho
Delaware
Vermont
Hawaii
Alaska
Ohio
Montana
Connecticut
New Mexico
Wisconsin
New Jersey
New Hampshire
Michigan
Louisiana
Illinois
Indiana
Minnesota
Kentucky
Kansas
Rhode Island
Oregon
Massachusetts
Colorado
Arizona
North Carolina
New York
Maine
Virginia
Tennessee
Mississippi
Maryland
Nevada
Missouri
Alabama
Washington
Pennsylvania
California
Georgia
South Carolina
Florida
Texas
North Dakota
Wyoming
South Dakota
Oklahoma
Arkansas
Utah
Iowa
Nebraska
0
20
40
60
Days since tenth case
Stay At Home Order
Not-Enacted
Enacted

Single and Nested Target Trials

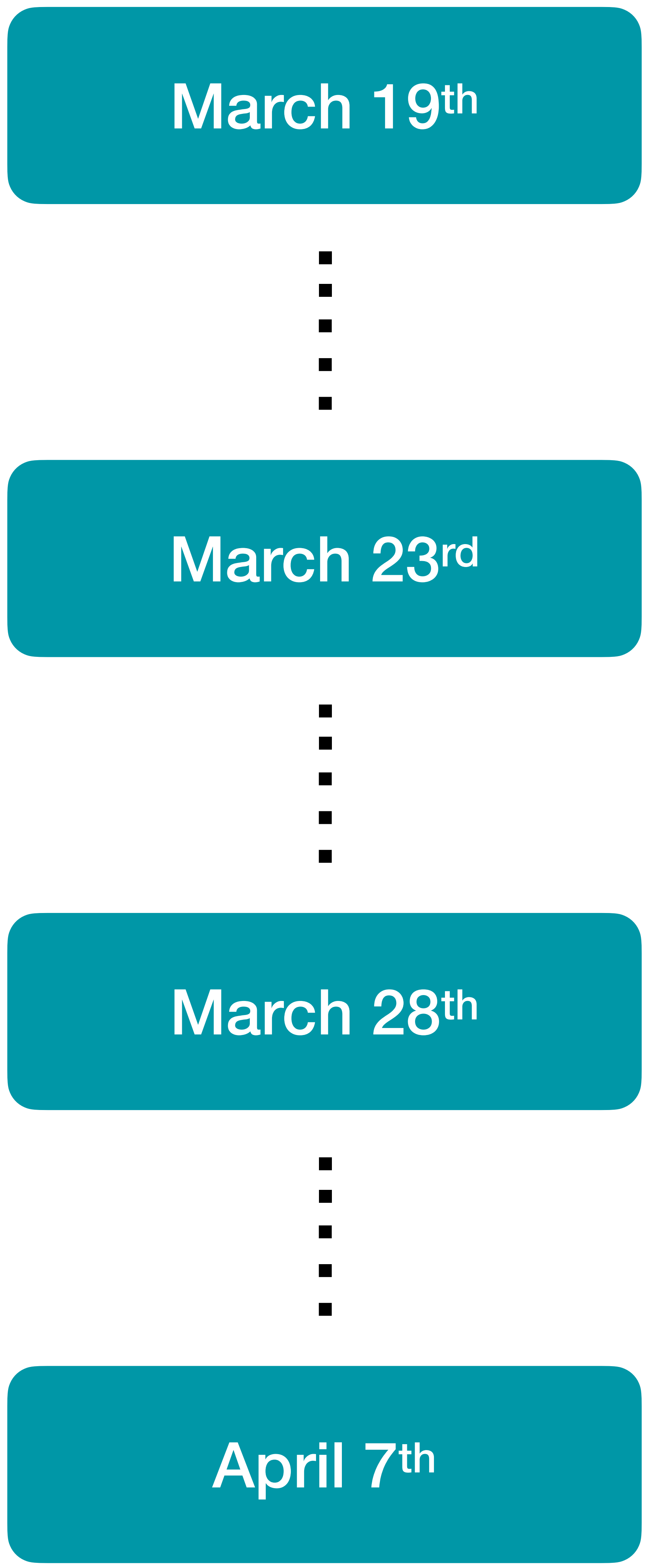
Target trials

Target trials

- Staggered adoption of policies

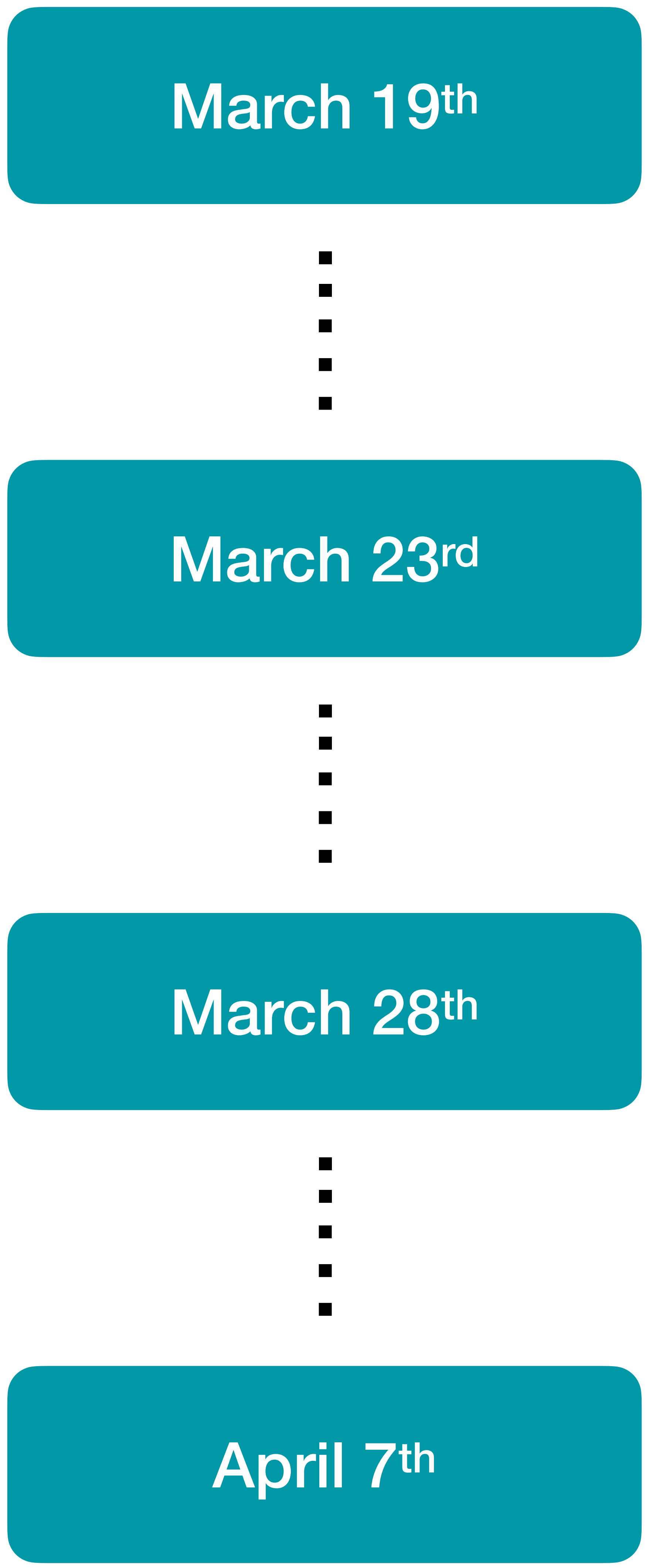
Target trials

- Staggered adoption of policies
- Create cohorts by adoption date



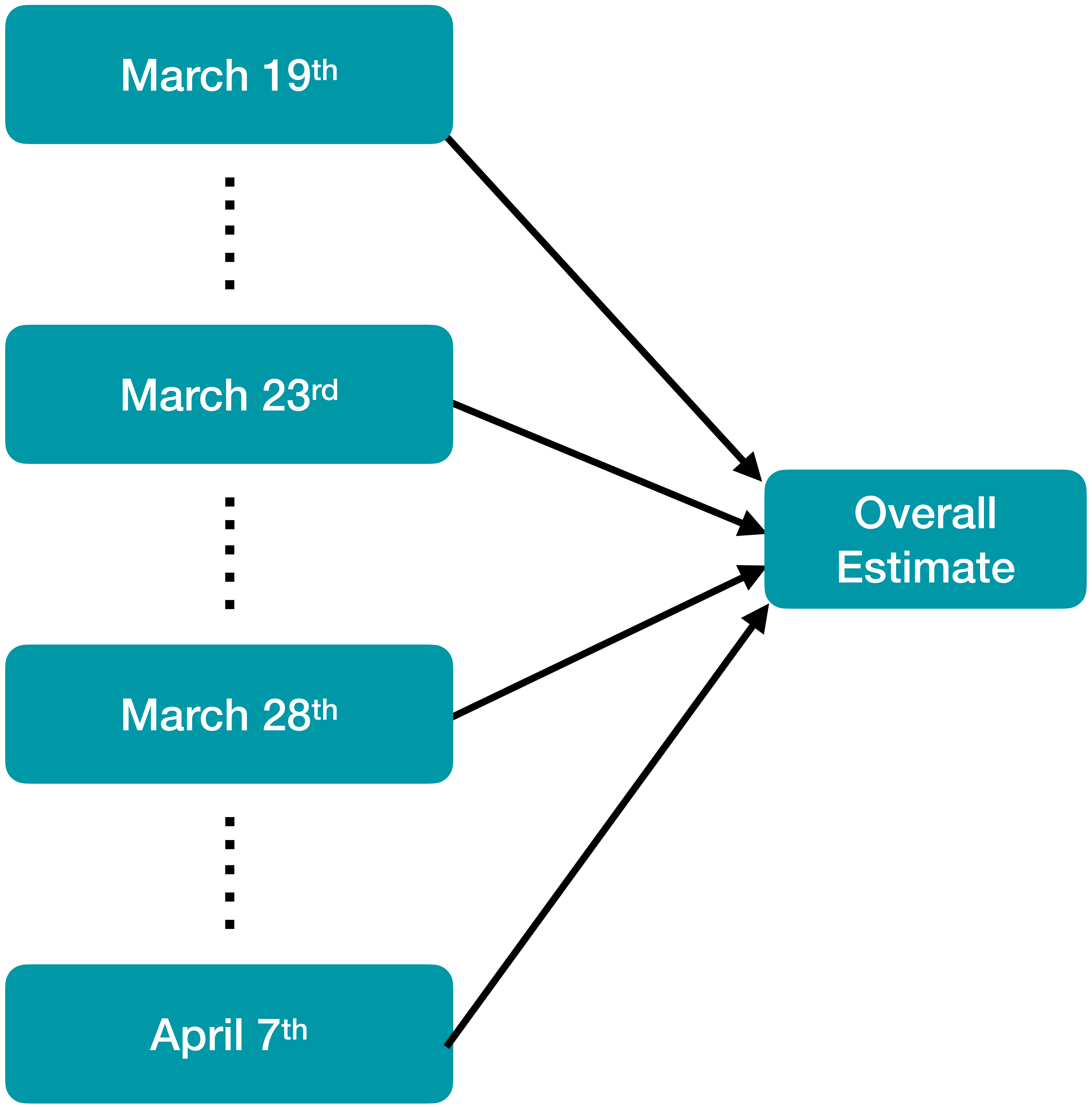
Target trials

- Staggered adoption of policies
- Create cohorts by adoption date
- Measure the effect for each cohort



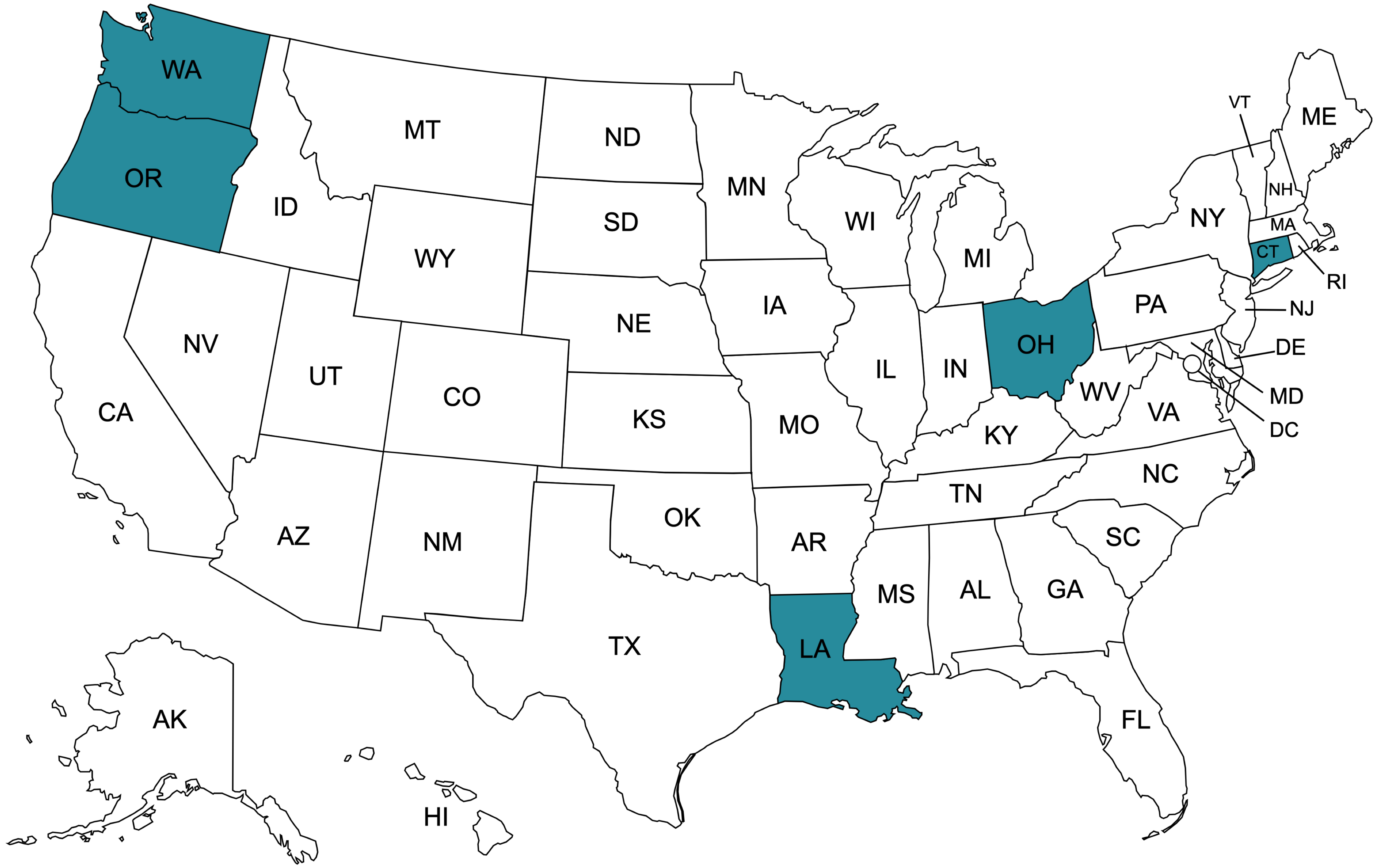
Target trials

- Staggered adoption of policies
- Create cohorts by adoption date
- Measure the effect for each cohort
- Aggregate across “single” target trials to a “nested” target trial



A single target trial

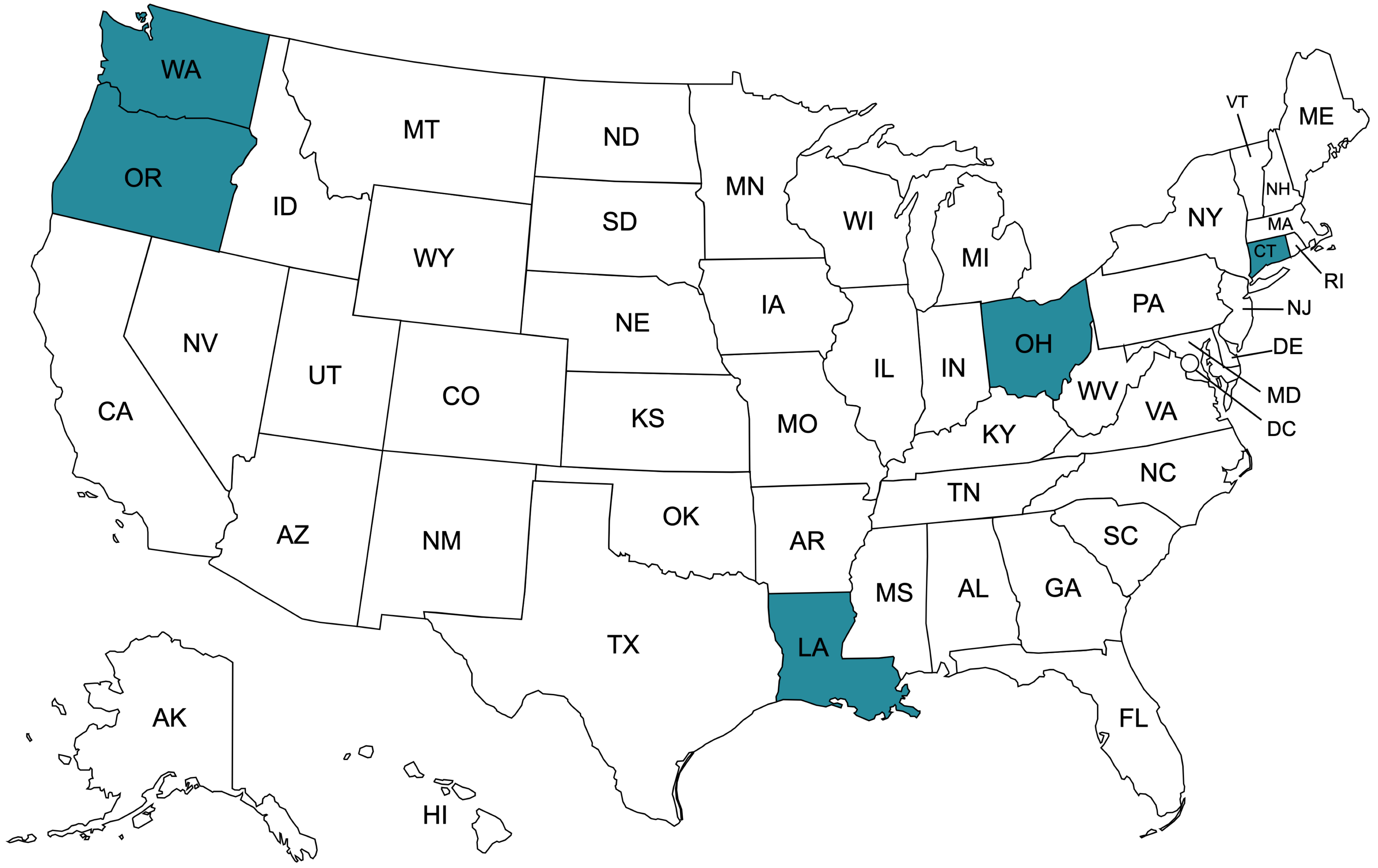
March 23rd Cohort



A single target trial

March 23rd Cohort

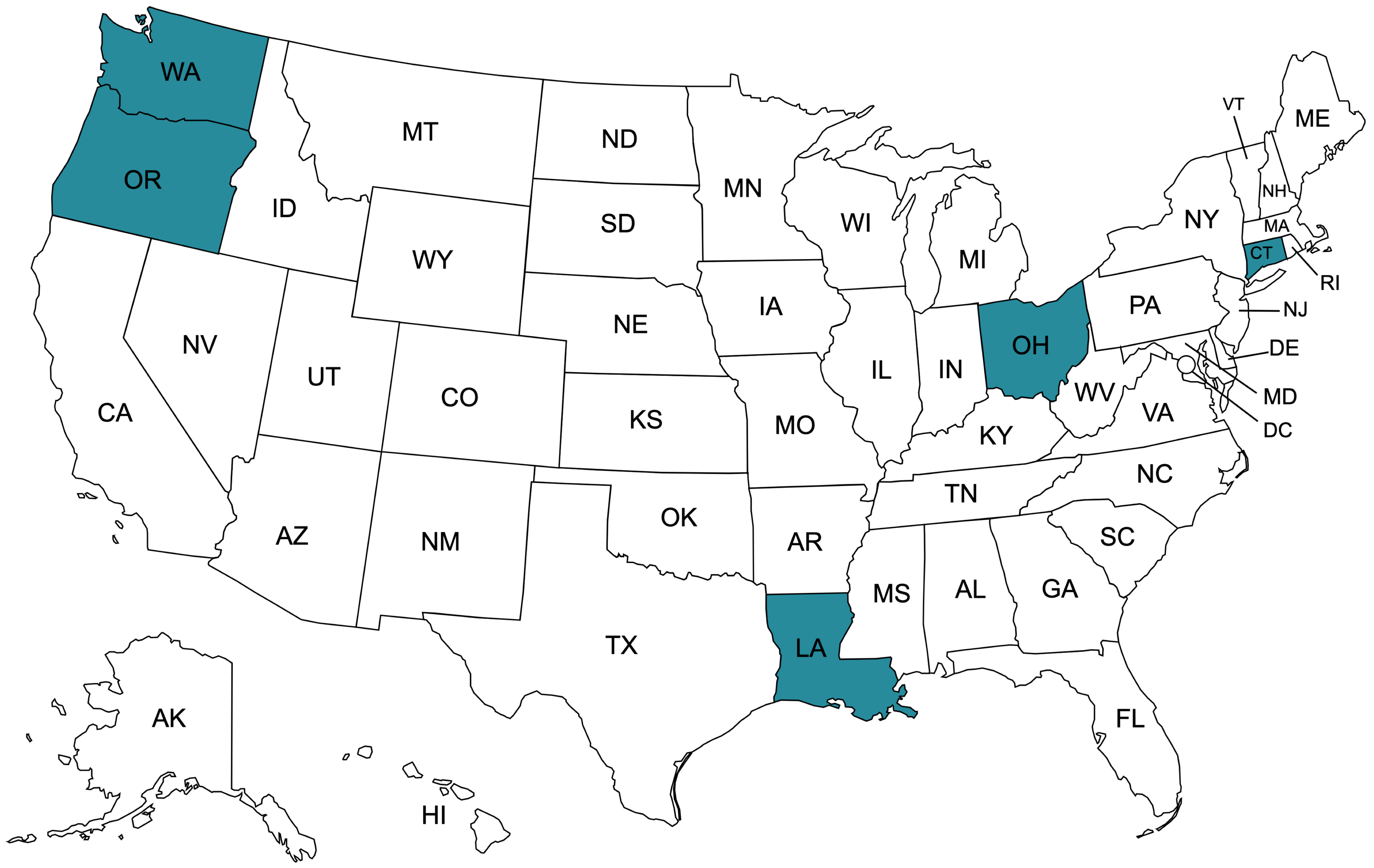
- What comparison states do we use?



A single target trial

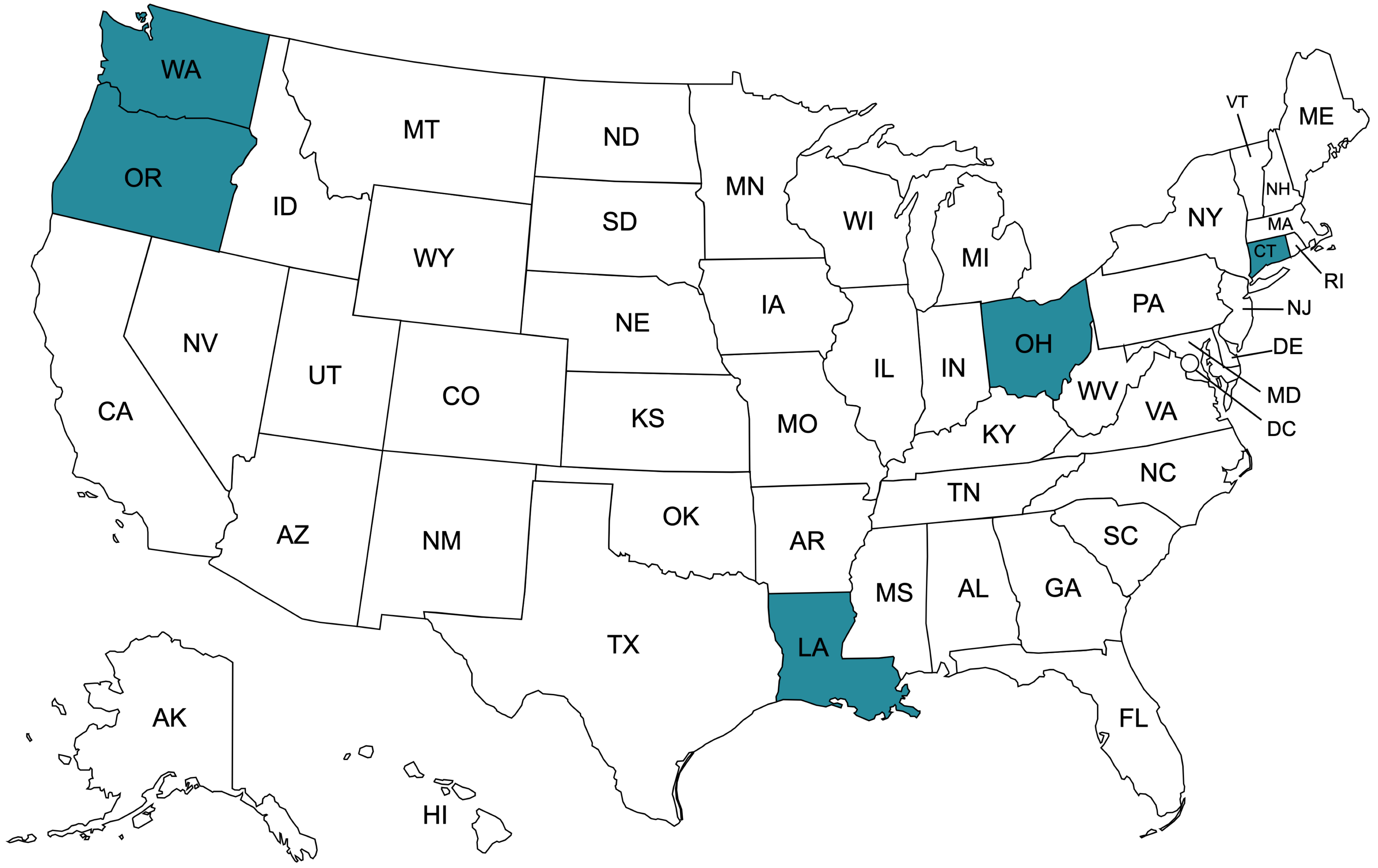
March 23rd Cohort

- What comparison states do we use?
- Length of follow up
  - Only 19 days between first and last adopters
  - Expect effects to be delayed



A single target trial
March 23rd Cohort

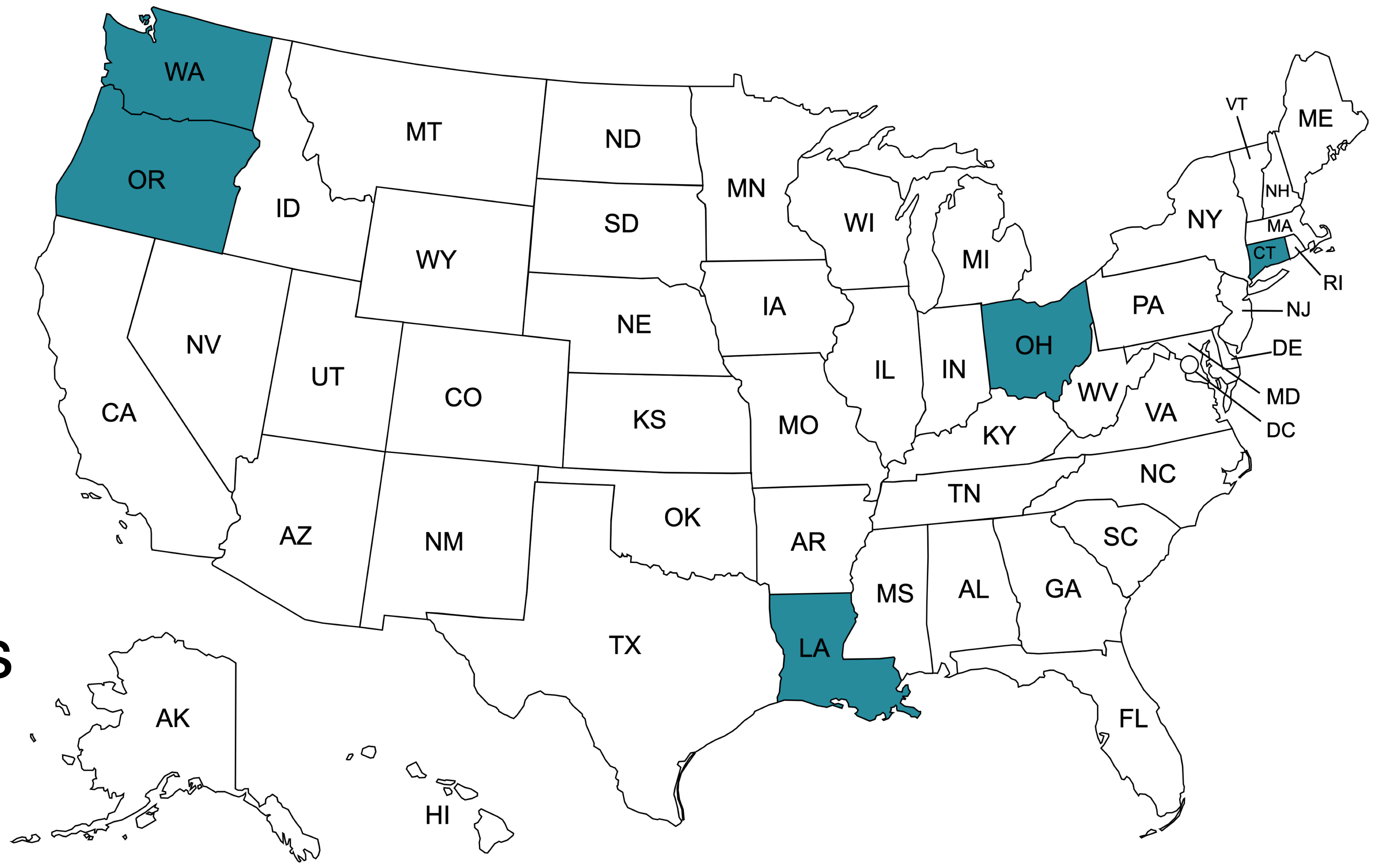
- What comparison states do we use?
- Length of follow up
  - Only 19 days between first and last adopters
  - Expect effects to be delayed
- Compare to 8 *never treated* states



A single target trial

March 23rd Cohort

- What comparison states do we use?
- Length of follow up
  - Only 19 days between first and last adopters
  - Expect effects to be delayed
- Compare to 8 *never treated* states
- Alternative: dynamic comparison groups
  - Need to assess assumptions for all groups
  - Are changes in effects just changes in comparison group?

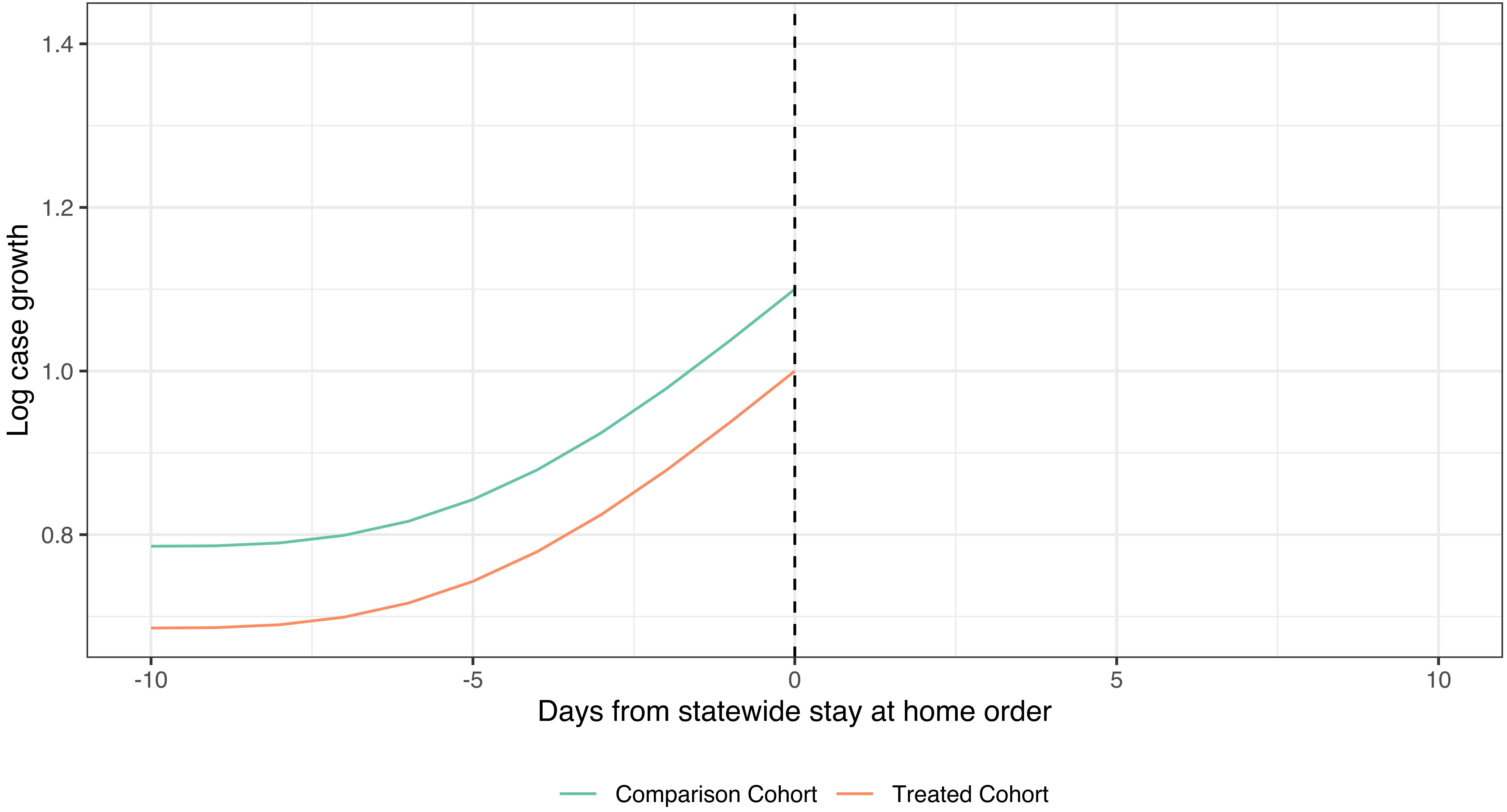


Analysis: Differences in Differences

Key assumption: parallel trends

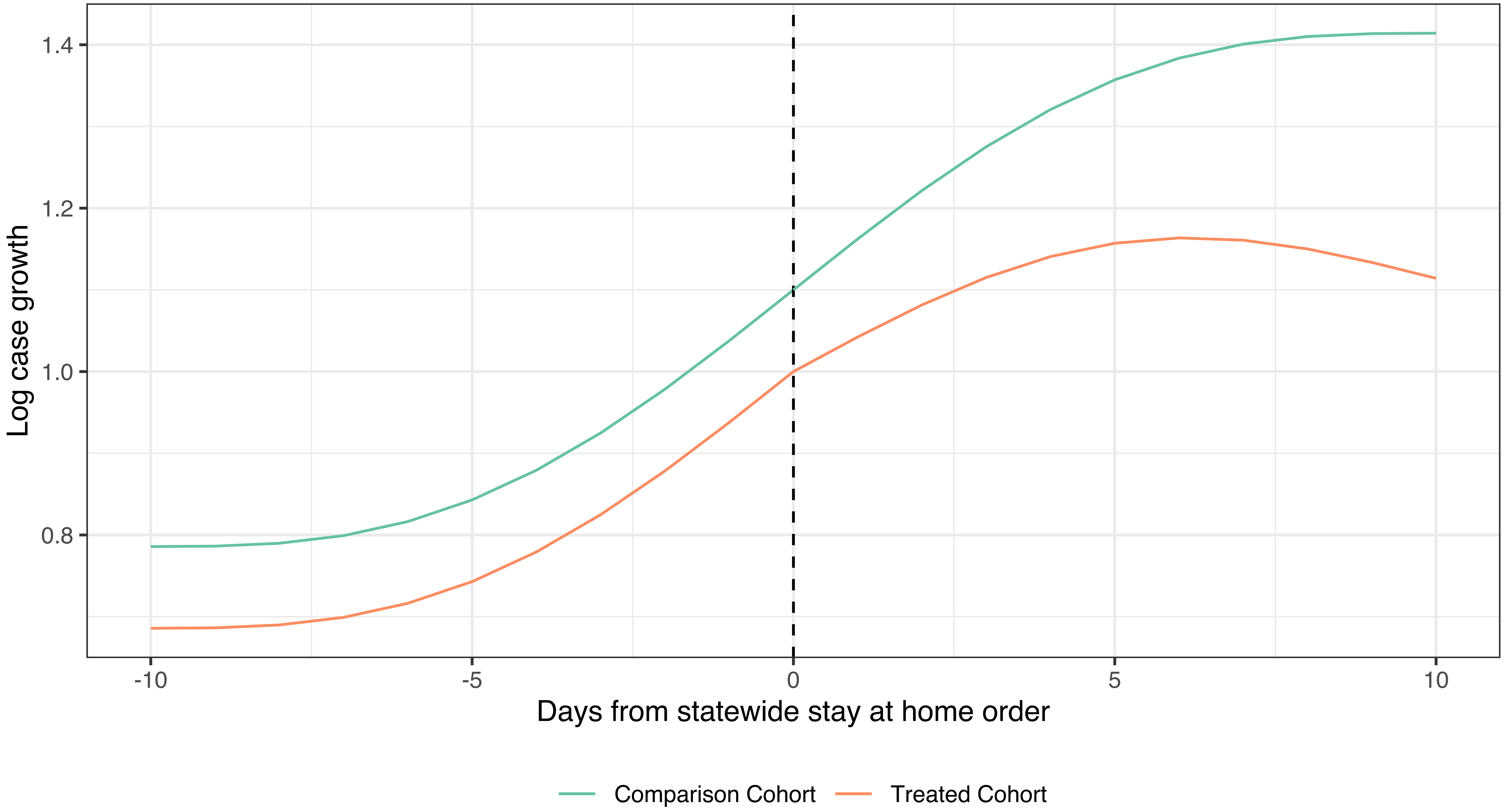
Analysis: Differences in Differences

Key assumption: parallel trends



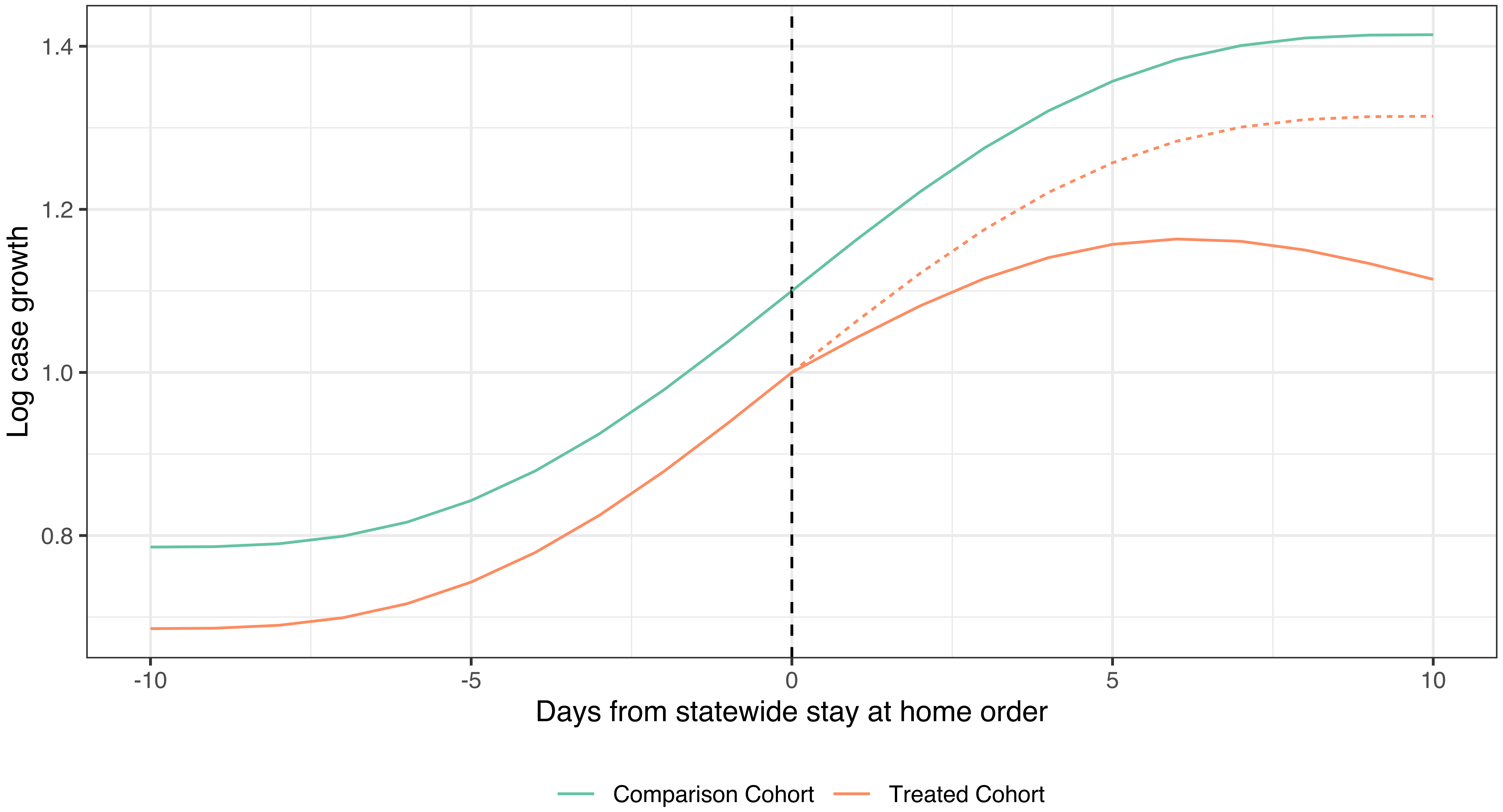
Analysis: Differences in Differences

Key assumption: parallel trends



Analysis: Differences in Differences

Key assumption: parallel trends



Estimating effects in a single target trial

Estimating effects in a single target trial

	Stay-at-Home Order		*Difference*
	Pre	**Post**	

Estimating effects in a single target trial

	Stay-at-Home Order		*Difference*
	Pre	Post	
March 23 Cohort	0.31 (37%)	0.09 (10%)	*-0.22 (-20%)*

Estimating effects in a single target trial

	Stay-at-Home Order		*Difference*
	Pre	**Post**	
March 23 Cohort	0.31 (37%)	0.09 (10%)	*-0.22 (-20%)*
Never Treated Cohort	0.24 (27%)	0.10 (11%)	*-0.14 (-12%)*

Estimating effects in a single target trial

	Stay-at-Home Order		*Difference*
	Pre	Post	
March 23 Cohort	0.31 (37%)	0.09 (10%)	*-0.22 (-20%)*
Never Treated Cohort	0.24 (27%)	0.10 (11%)	*-0.14 (-12%)*
Difference	*+0.07 (+10%)*	*-0.01 (-1%)*	***-0.08 (-8%)***

Estimating effects in a single target trial

	Stay-at-Home Order		*Difference*
	Pre	**Post**	
March 23 Cohort	0.31 (37%)	0.09 (10%)	*-0.22 (-20%)*
Never Treated Cohort	0.24 (27%)	0.10 (11%)	*-0.14 (-12%)*
Difference	*+0.07 (+10%)*	*-0.01 (-1%)*	***-0.08 (-8%)***

Estimating time varying effects

Estimating time varying effects

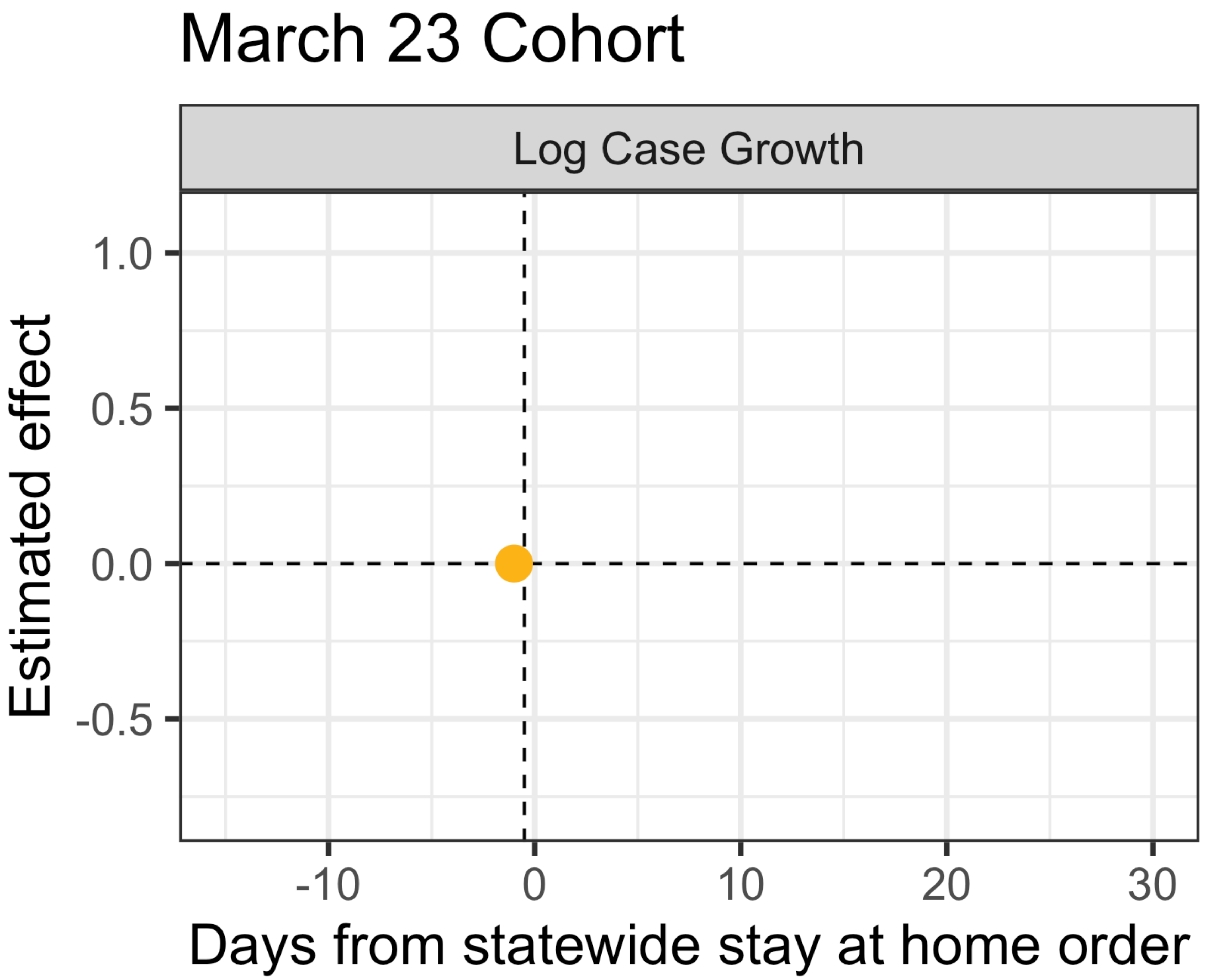
- 2x2 DiD table is blunt
  - Averages over the entire post period

Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as “pre” period
  - Create a sequence of 2x2 estimates

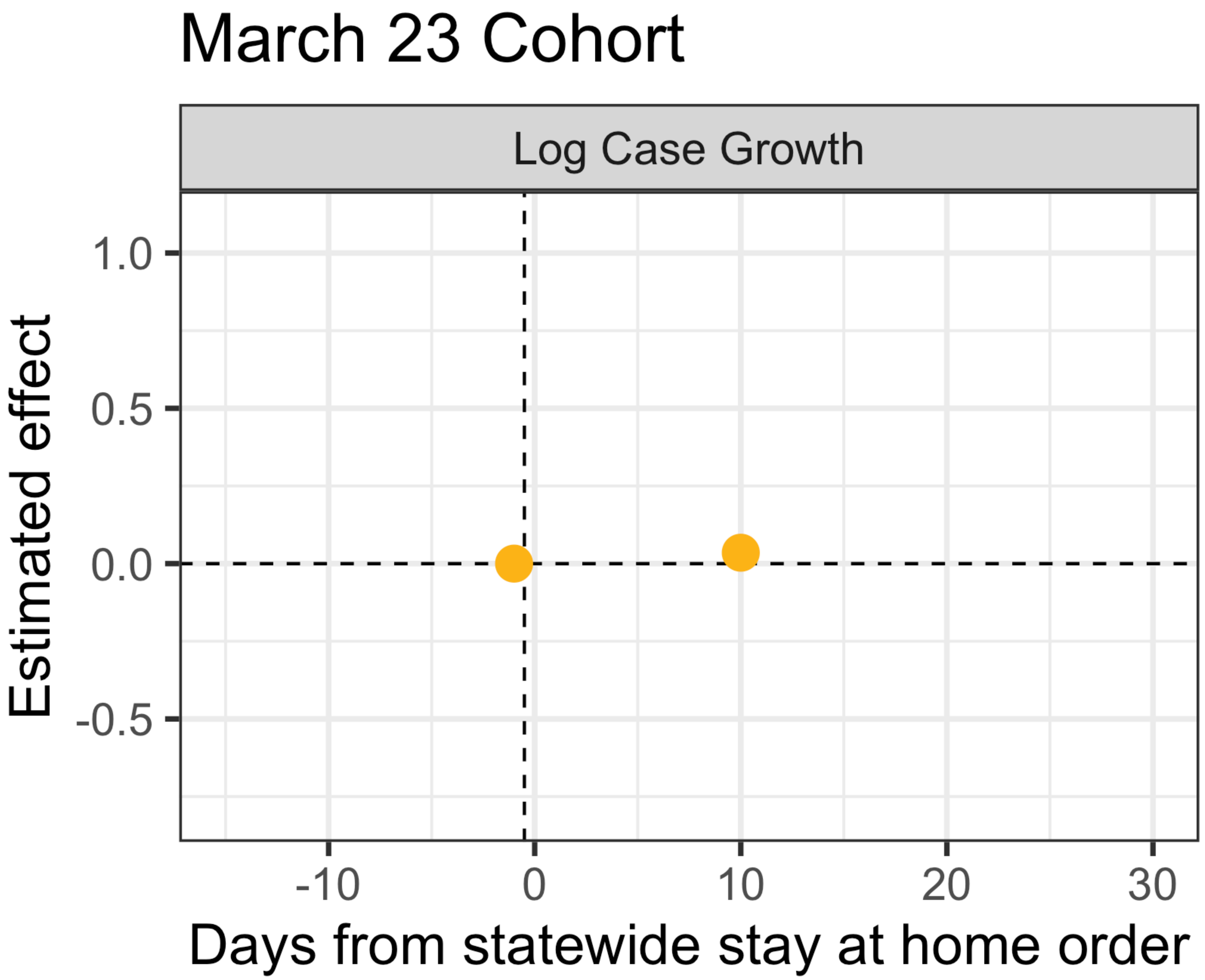
Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as “pre” period
  - Create a sequence of 2x2 estimates



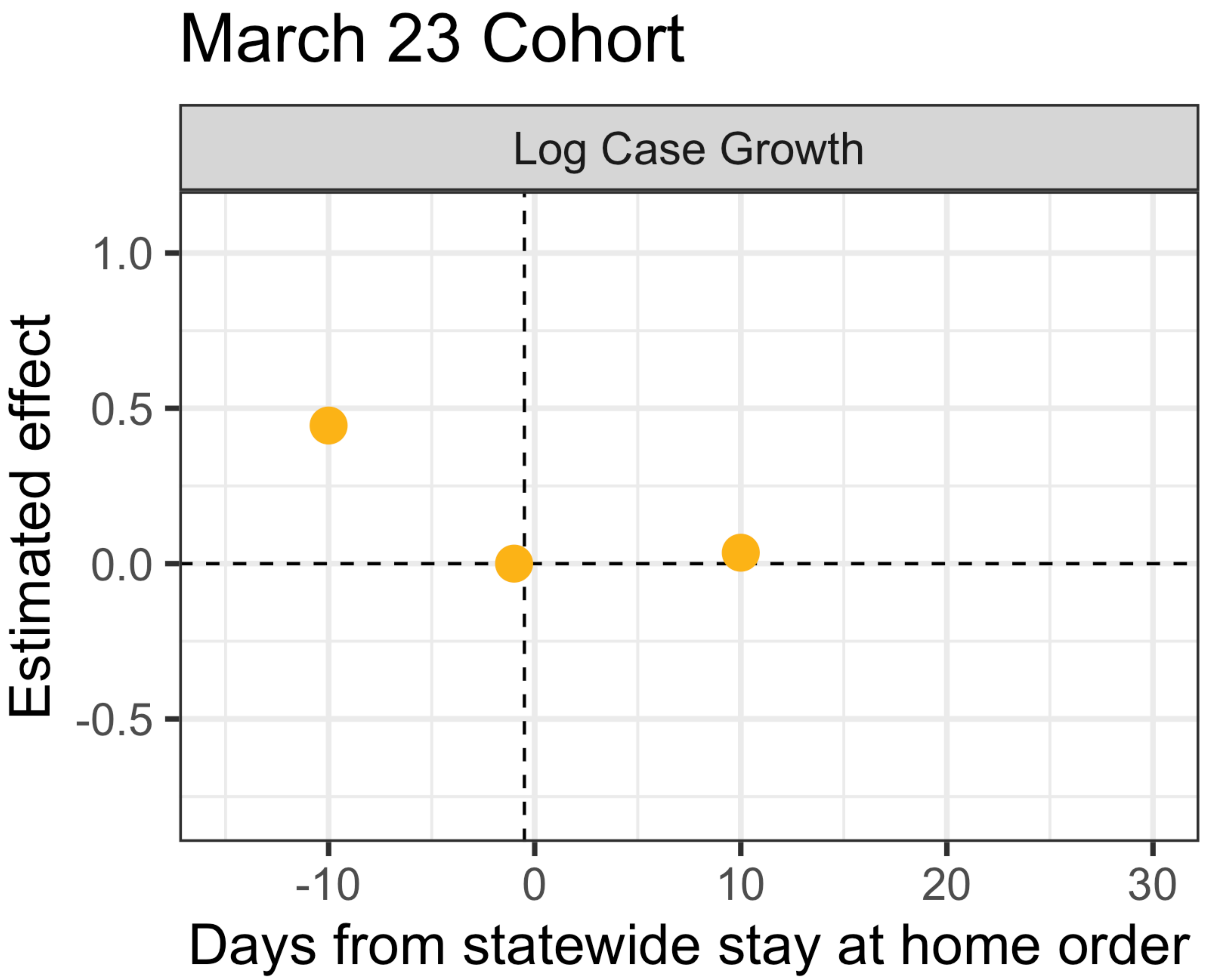
Estimating time varying effects

- 2x2 DiD table is blunt
    - Averages over the entire post period
- Dynamic DiD estimate
    - Use one reference time as “pre” period
    - Create a sequence of 2x2 estimates



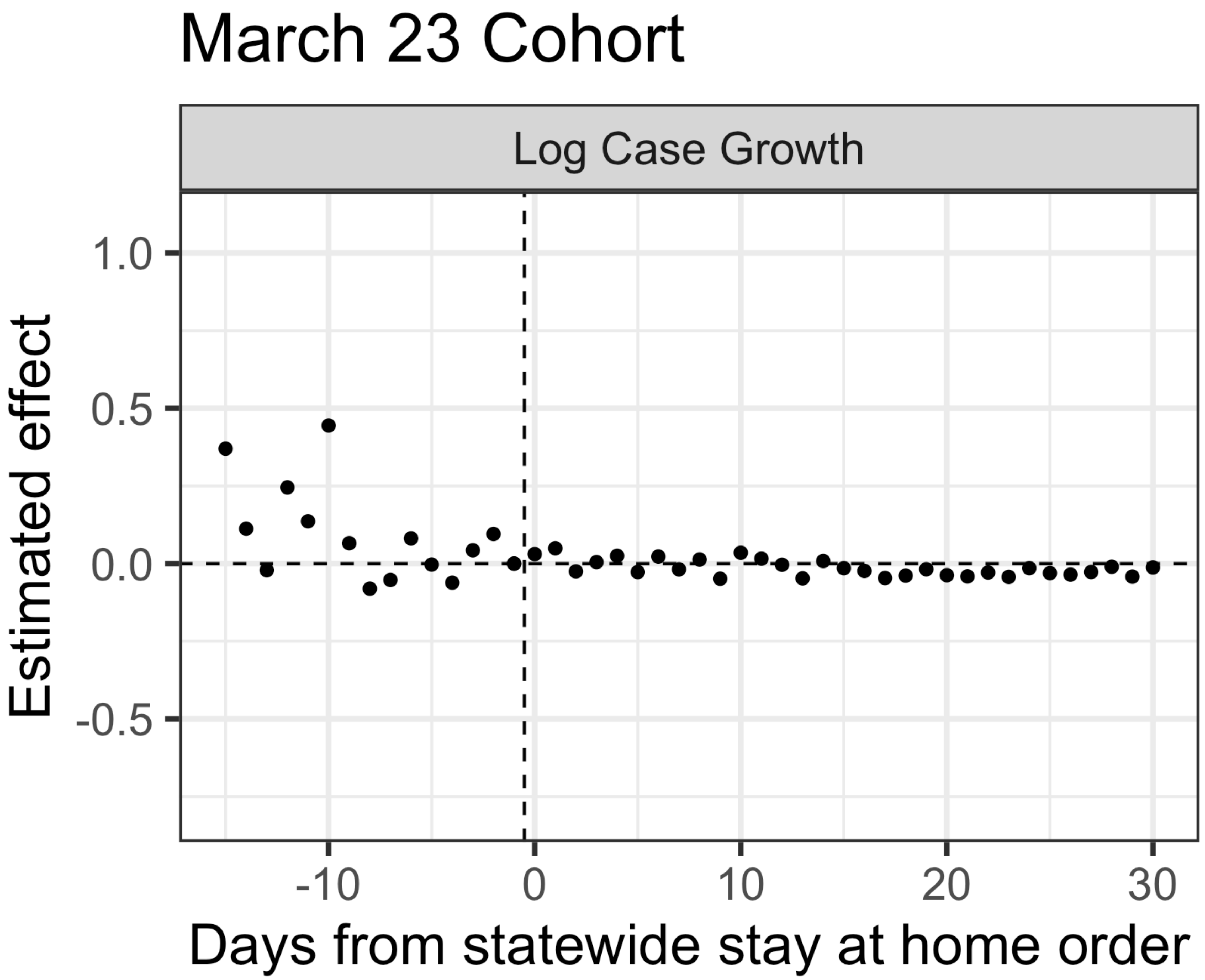
Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as “pre” period
  - Create a sequence of 2x2 estimates



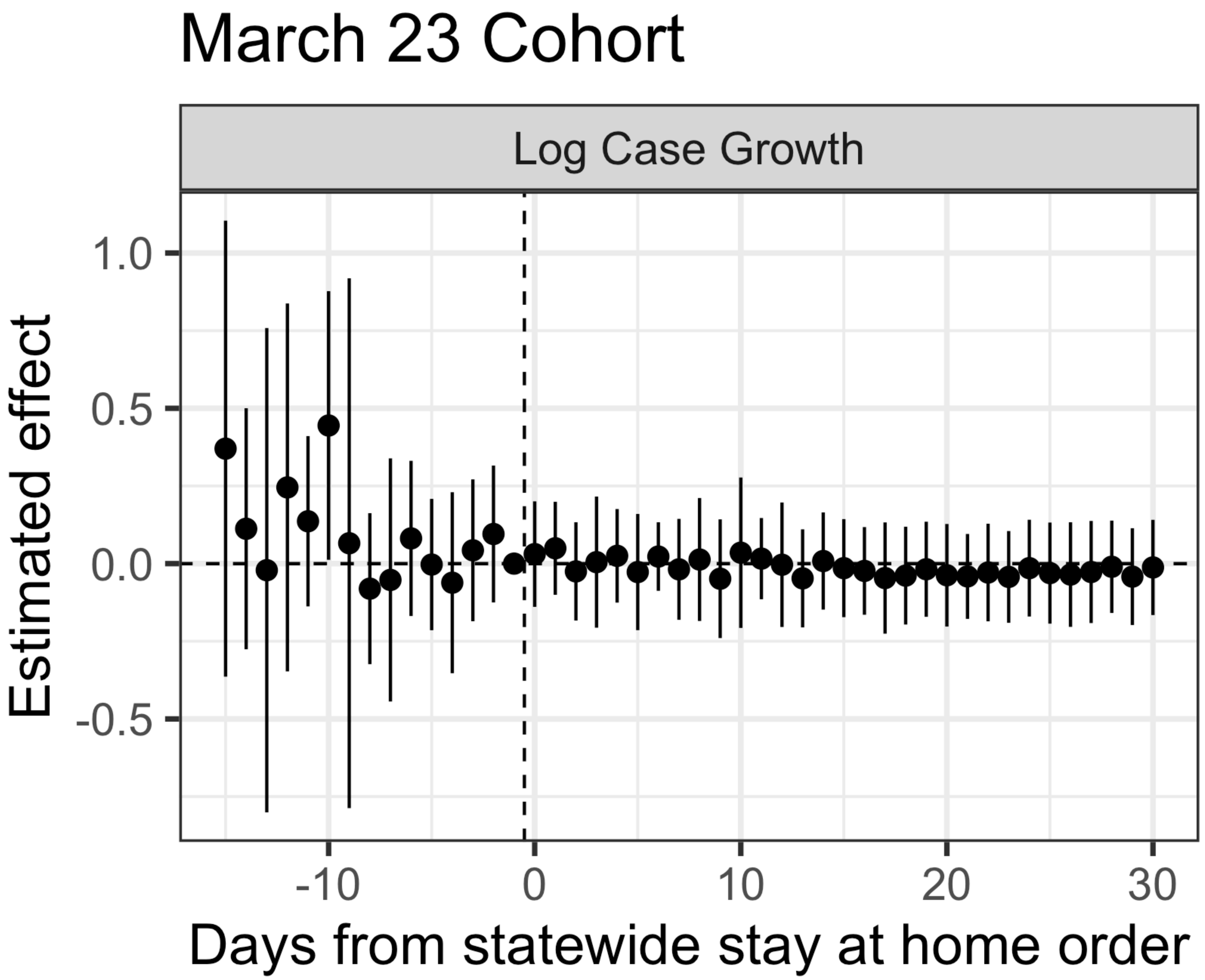
Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as "pre" period
  - Create a sequence of 2x2 estimates



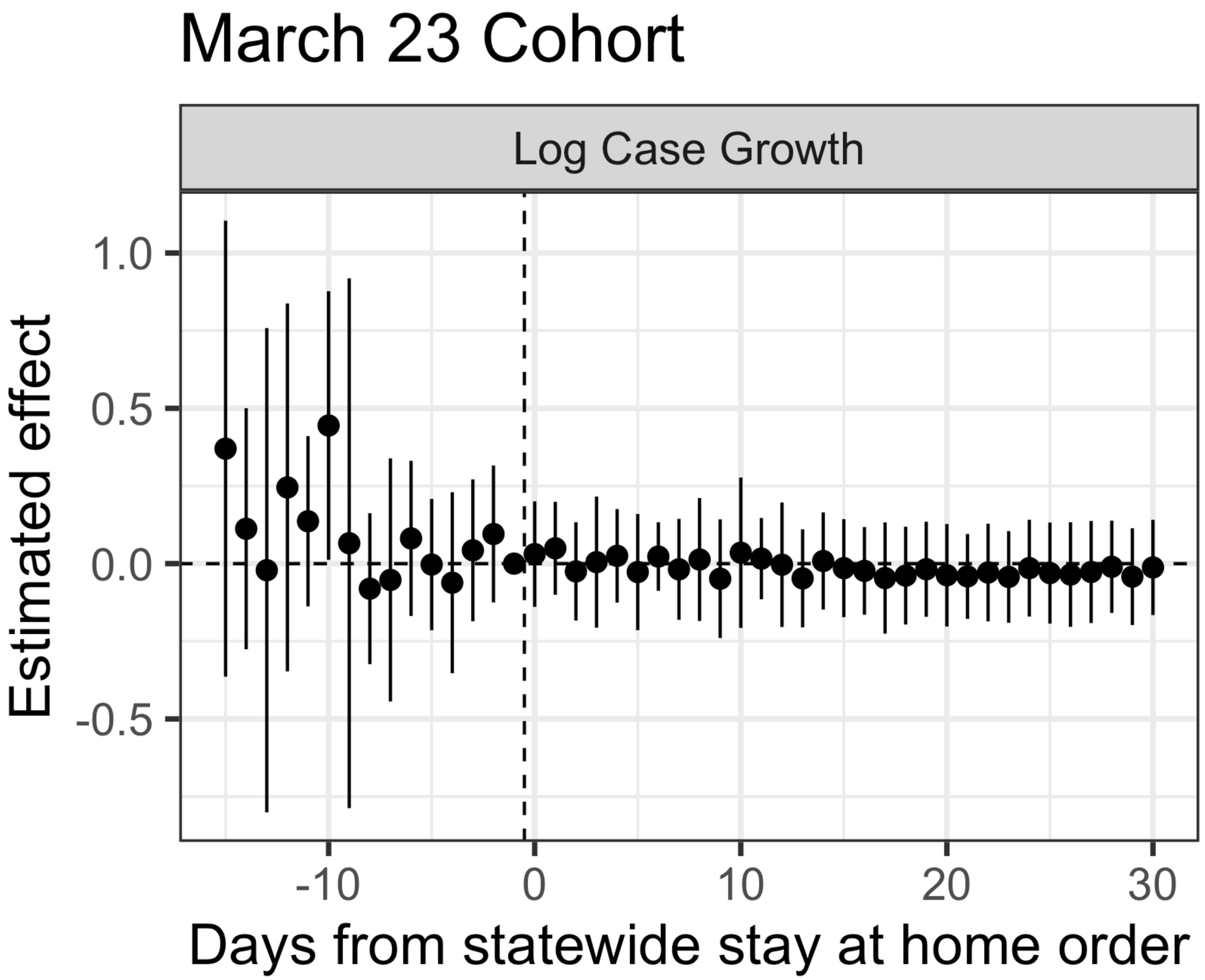
Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as “pre” period
  - Create a sequence of 2x2 estimates



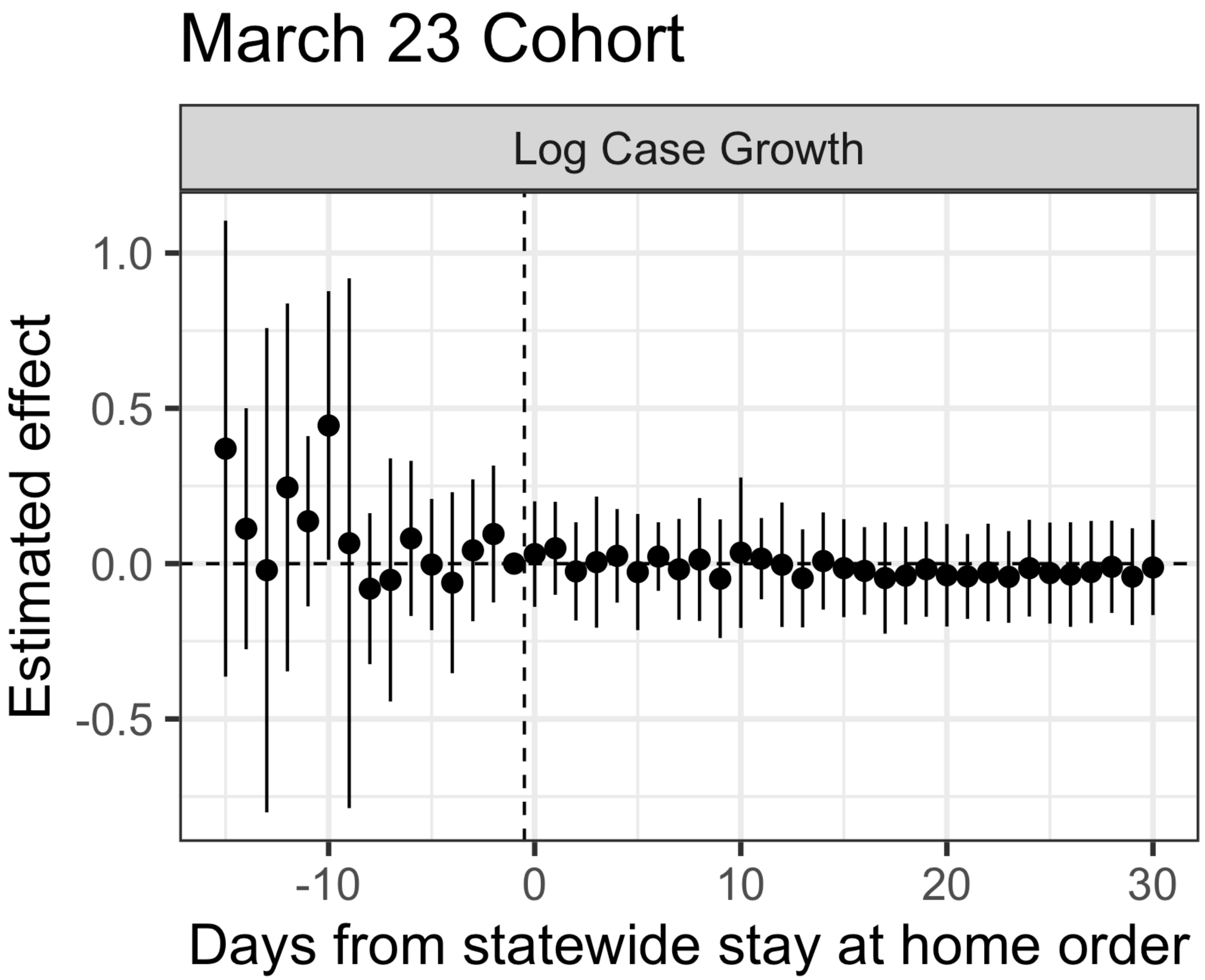
Estimating time varying effects

- 2x2 DiD table is blunt
    - Averages over the entire post period
- Dynamic DiD estimate
    - Use one reference time as “pre” period
    - Create a sequence of 2x2 estimates
- Allows for a diagnostic check
    - Pre-period should have **zero** effect



Estimating time varying effects

- 2x2 DiD table is blunt
  - Averages over the entire post period
- Dynamic DiD estimate
  - Use one reference time as “pre” period
  - Create a sequence of 2x2 estimates
- Allows for a diagnostic check
  - Pre-period should have **zero** effect
- Possible violations of // trends
  - Anticipation, time-varying confounding

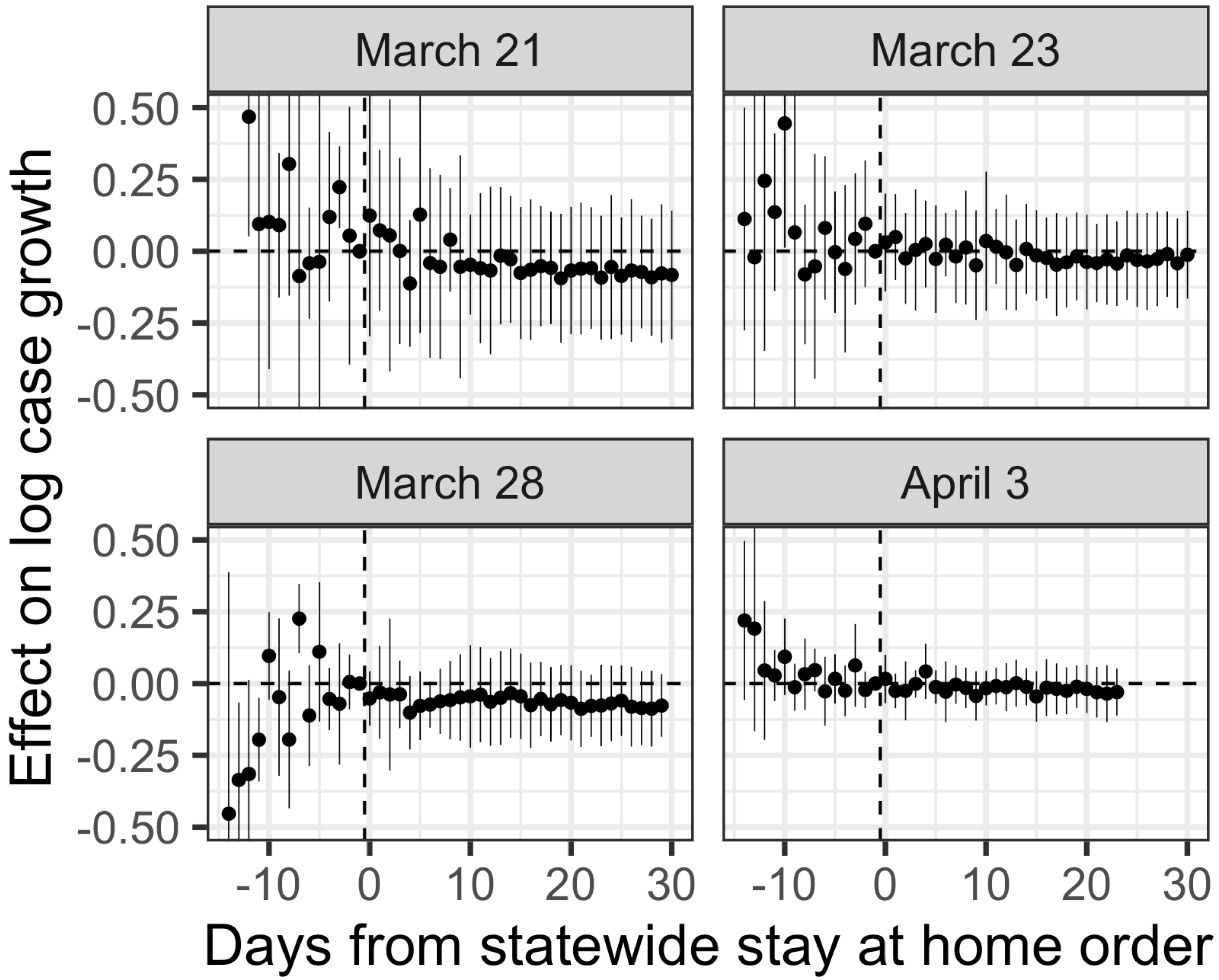


From single to nested target trials

From single to nested target trials

- Each single trial is different
  - Starting point, length of follow up, etc.

[Hernán et al 2016]



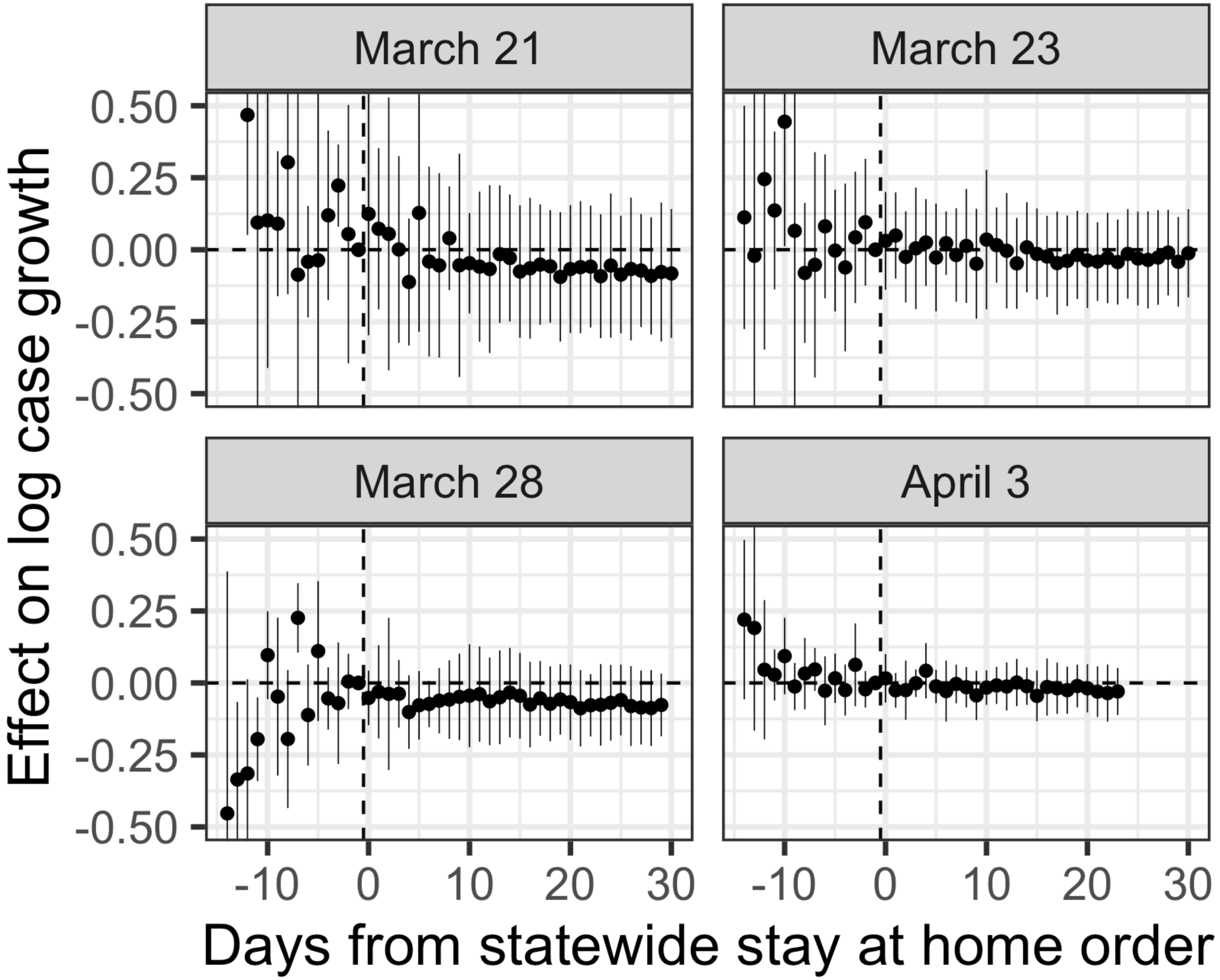
From single to nested target trials

- Each single trial is different
  - Starting point, length of follow up, etc.

[Hernán et al 2016]

- Take a size-weighted average

$$\widehat{DID}_k = \frac{1}{n_1}\sum_{g=1}^{G} n_{1g}\widehat{DID}_{kg}$$



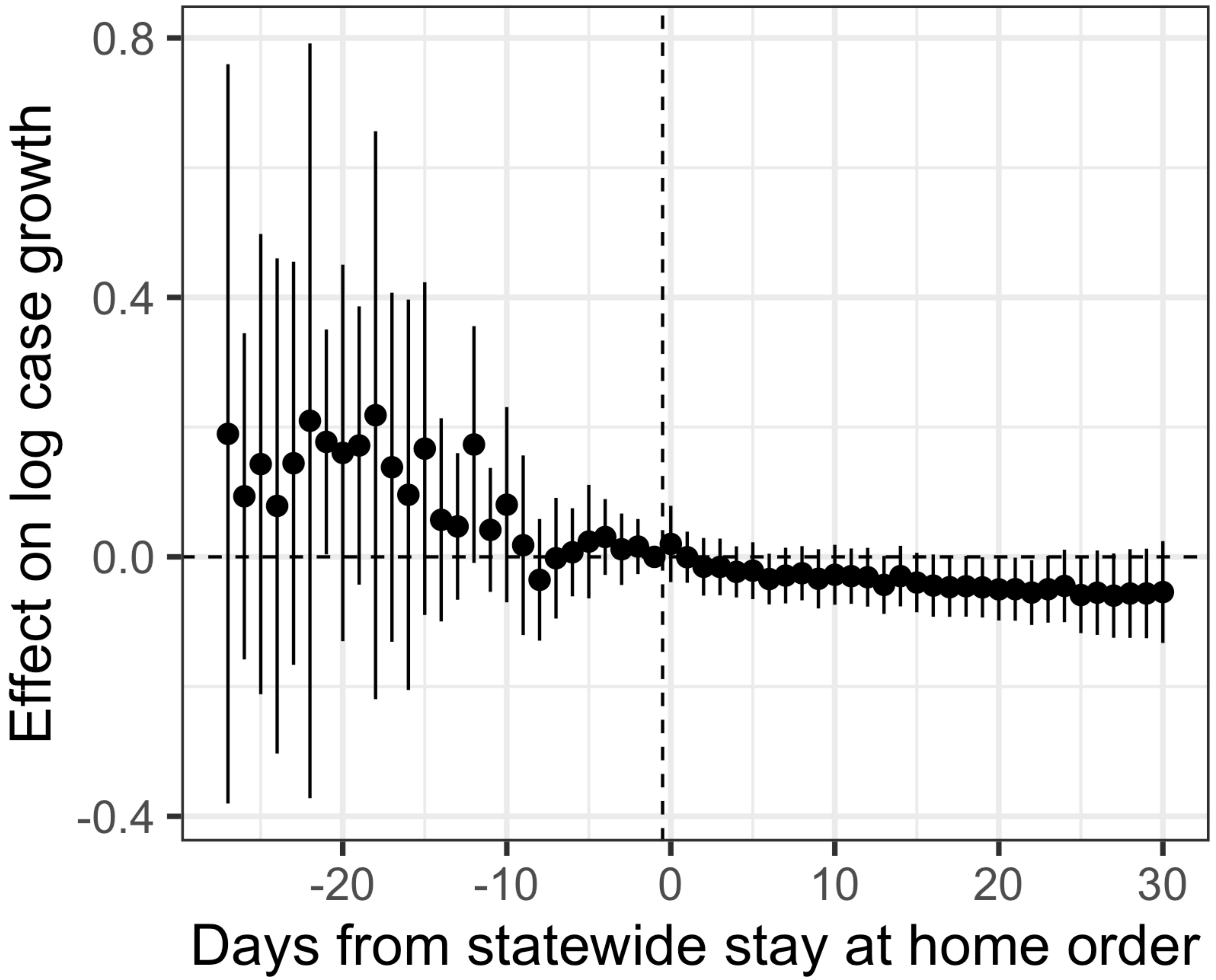
From single to nested target trials

- Each single trial is different
    - Starting point, length of follow up, etc.

[Hernán et al 2016]

- Take a size-weighted average

$$\widehat{DID}_k = \frac{1}{n_1} \sum_{g=1}^{G} n_{1g} \widehat{DID}_{kg}$$



From single to nested target trials

- Each single trial is different
  - Starting point, length of follow up, etc.

[Hernán et al 2016]

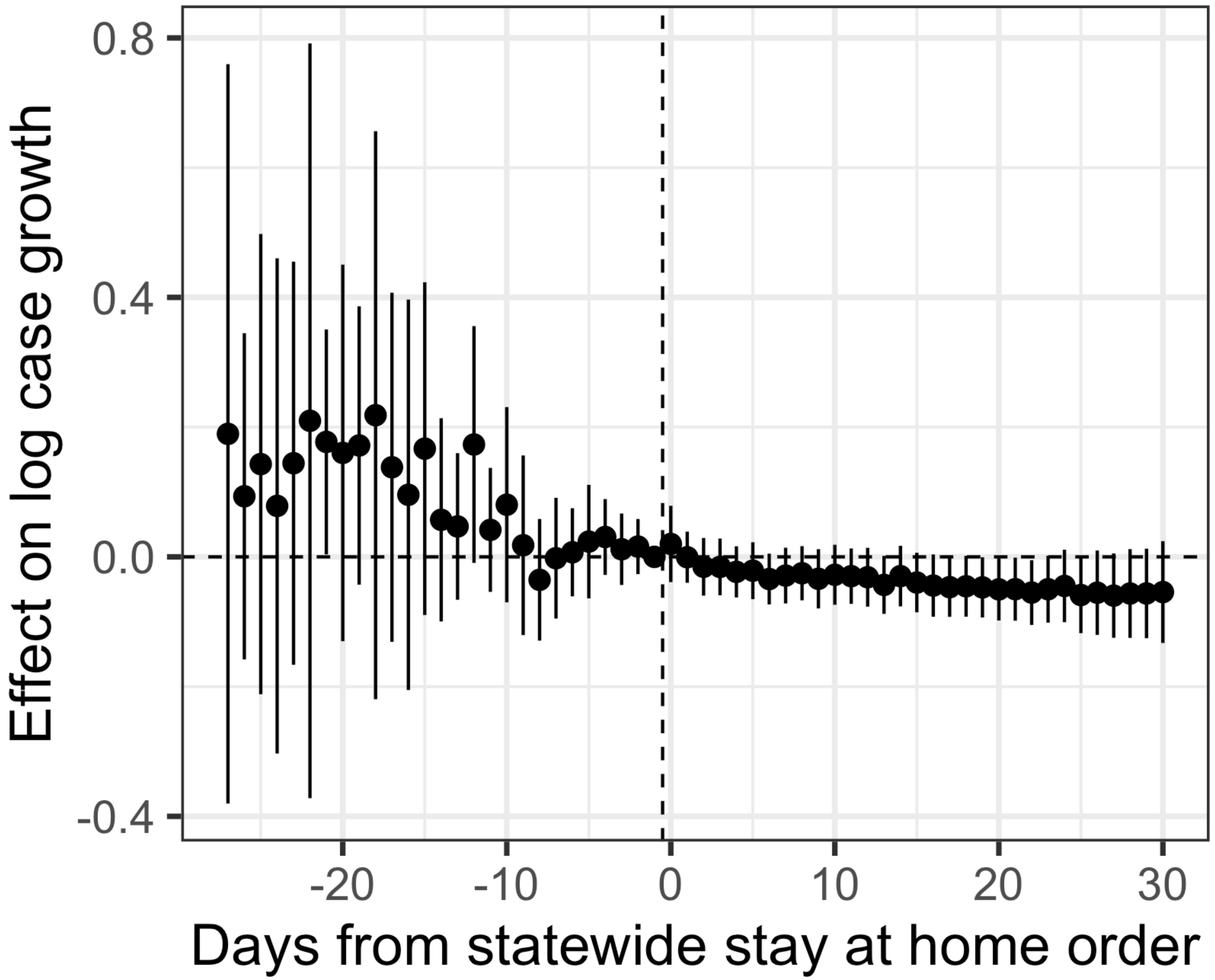
- Take a size-weighted average

$$\widehat{DID}_k = \frac{1}{n_1} \sum_{g=1}^{G} n_{1g} \widehat{DID}_{kg}$$

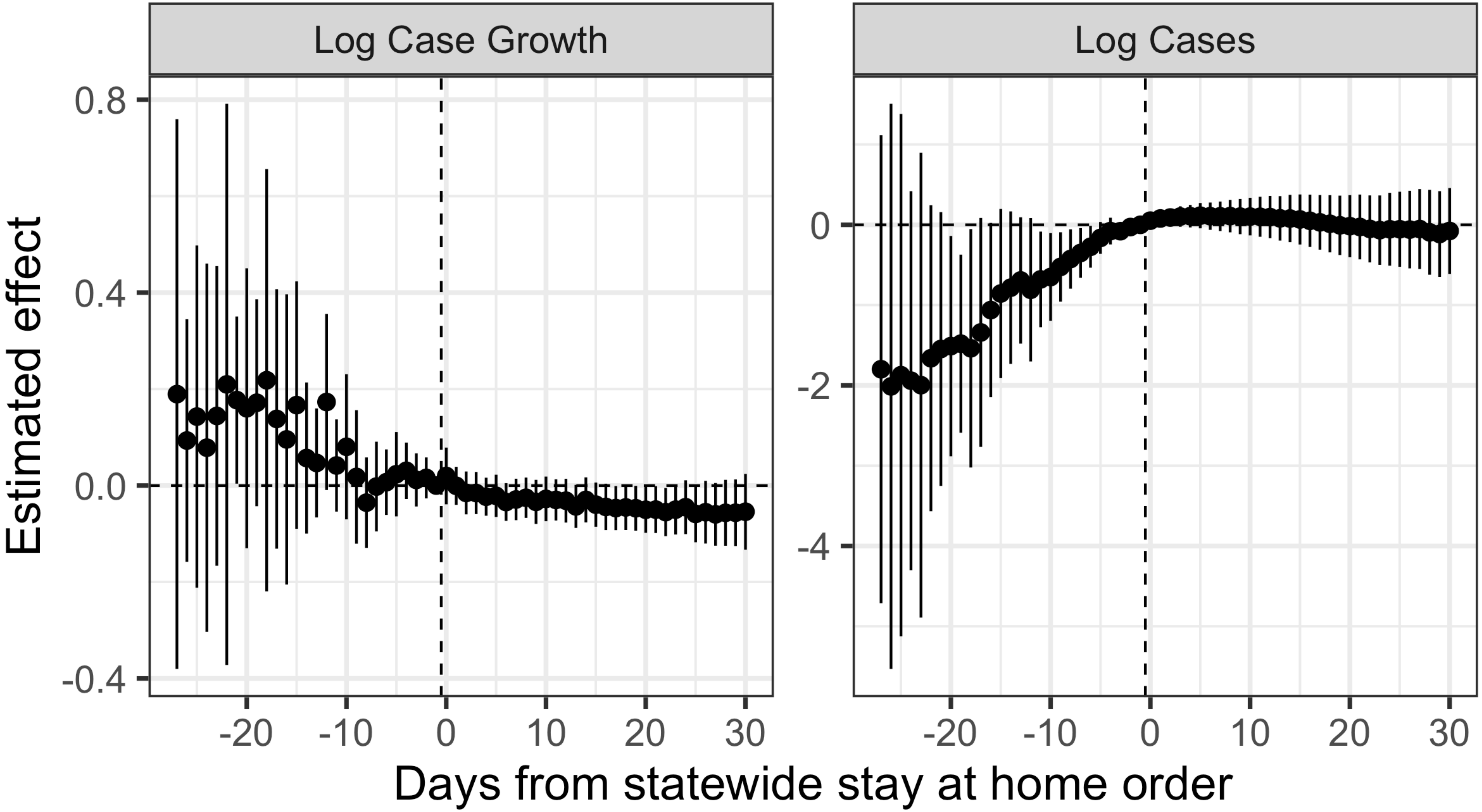
- Recovers “stacked” DiD

- Uncertainty quantification is tricky
  - Various forms of resampling methods

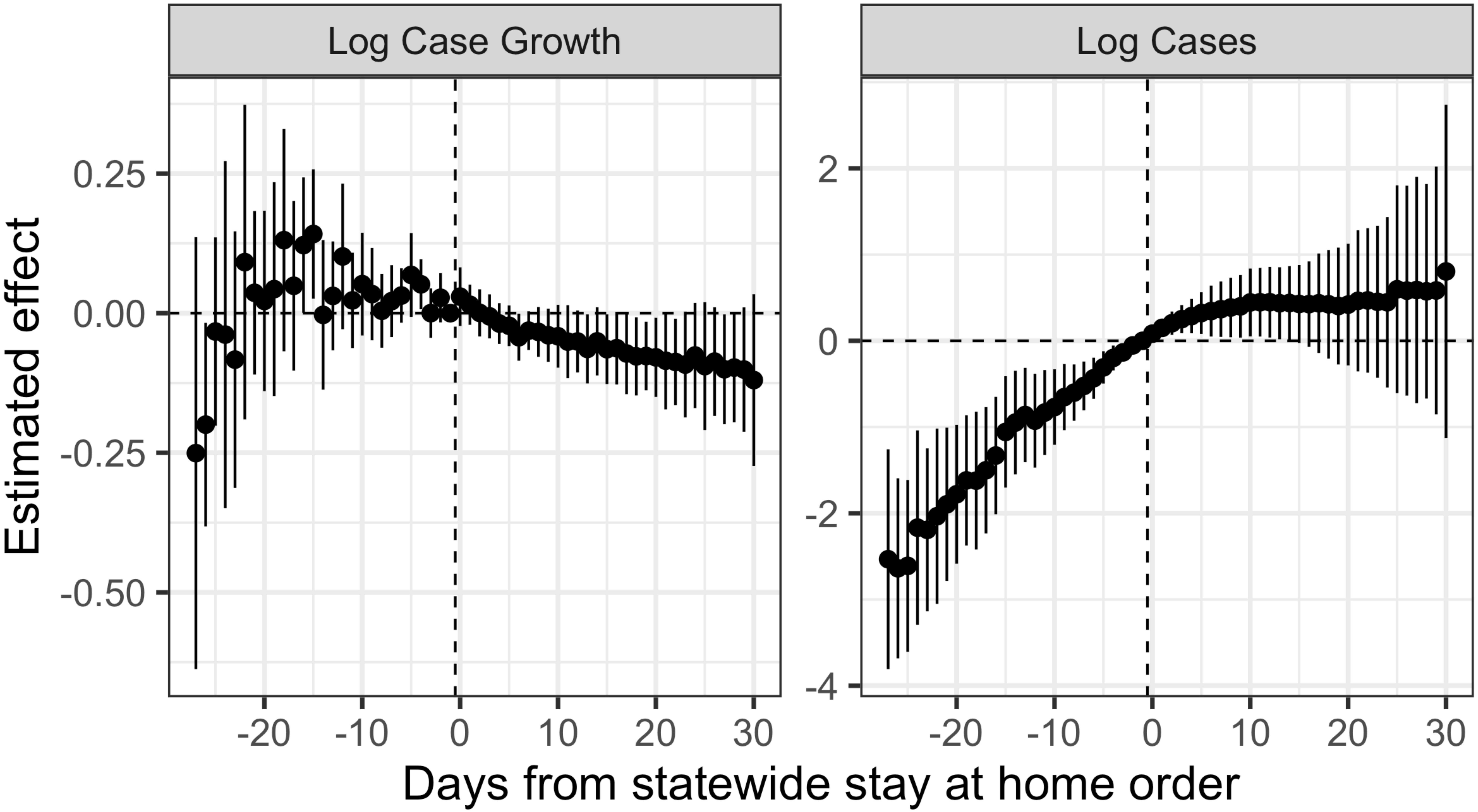
[Abraham & Sun 2021; Callaway & Sant’Anna 2021]



Plausibility of // trends depends on outcome



Slight differences when using case time



Wrapping up

Wrapping up

- Evaluating policy impact is difficult, especially recently!

Wrapping up

- Evaluating policy impact is difficult, especially recently!
- More design-based thinking can give some clarity
  - Avoids many of the pitfalls of naive regression models
  - Newer panel-data approaches fit naturally into a trial-emulation framework
  - IPW, matching, double robust DiD, synthetic controls, etc.
  - How we can use models to help estimate effects, without relying on them for identification

Wrapping up

- Evaluating policy impact is difficult, especially recently!
- More design-based thinking can give some clarity
  - Avoids many of the pitfalls of naive regression models
  - Newer panel-data approaches fit naturally into a trial-emulation framework
  - IPW, matching, double robust DiD, synthetic controls, etc.
  - How we can use models to help estimate effects, without relying on them for identification

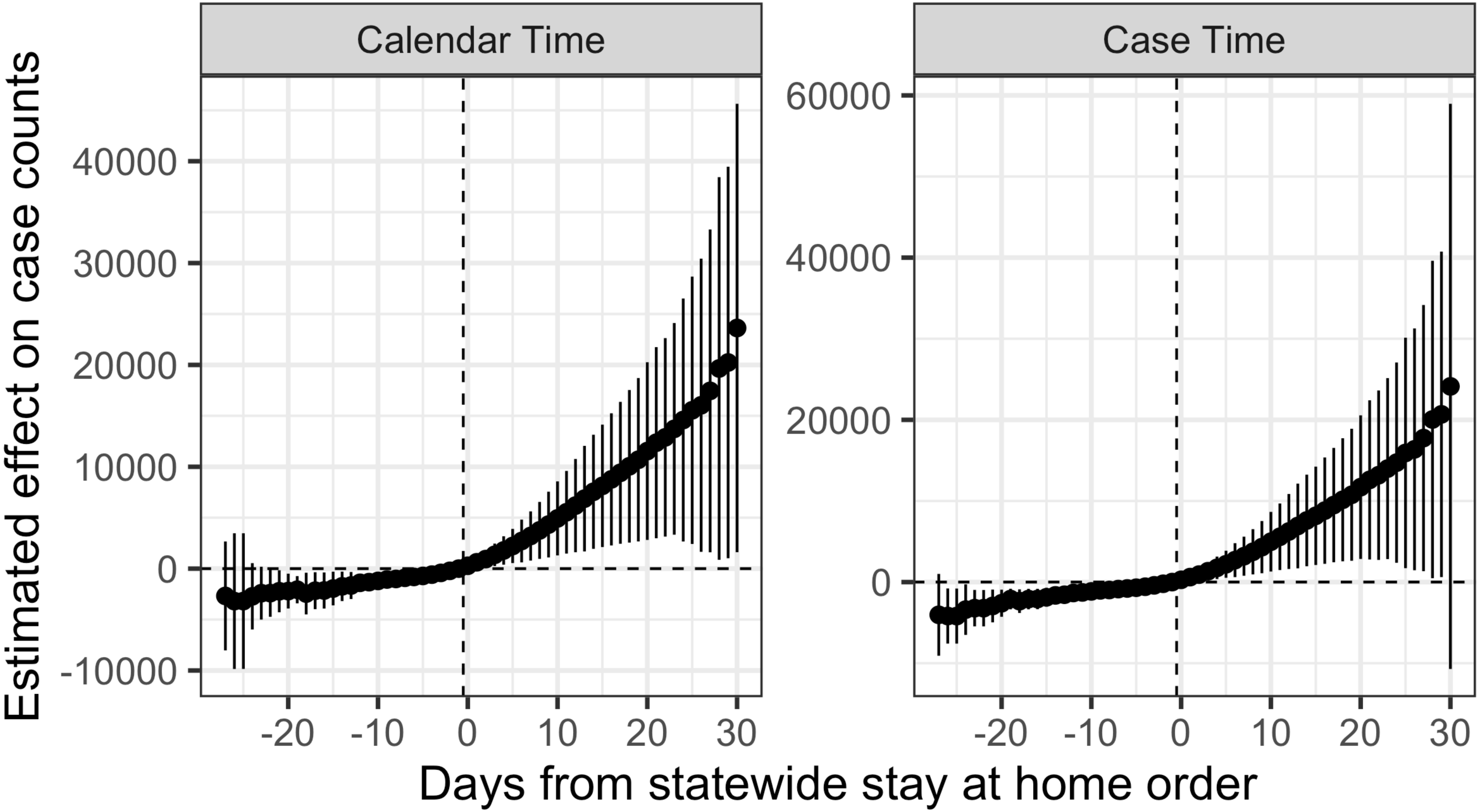
Thank you!

ebenmichael.github.io

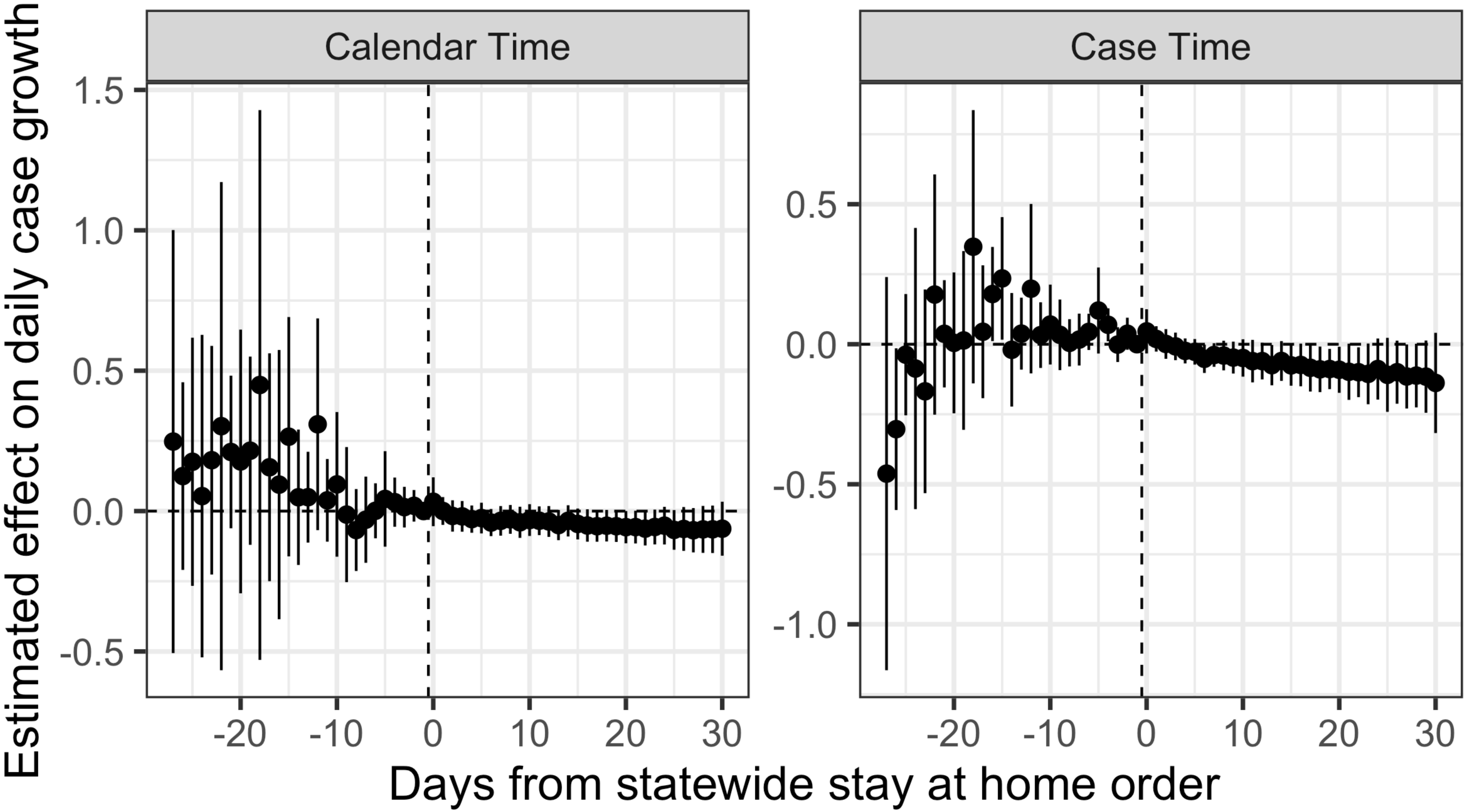
Ben-Michael, E., Feller, A., & Stuart, E. A. (2021). A Trial Emulation Approach for Policy Evaluations with Group-level Longitudinal Data. Epidemiology, 32(4), 533–540.

Appendix

Raw case count estimates



Case ratio estimates



References

- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. Journal of Econometrics, 225(2), 200–230.
- Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. Journal of Clinical Epidemiology. 2018;96:12-22. doi:10.1016/j.jclinepi.2017.11.02
- Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernan MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nature Medicine. 2019; 25, 1601–1606. https://doi.org/10.1038/s41591-019-0597-x [5][19]
- Goolsbee A, Syverson C. (2021) Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline. Journal of Public Economics, 193,
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of Clinical Epidemiology. 2016; 79, 70–75.
- The New York Times. Coronavirus (Covid-19) Data in the United States. Accessed August 2, 2020. https://github.com/nytimes/covid-19-data/blob/master/README.md