Living in a Parallel World?

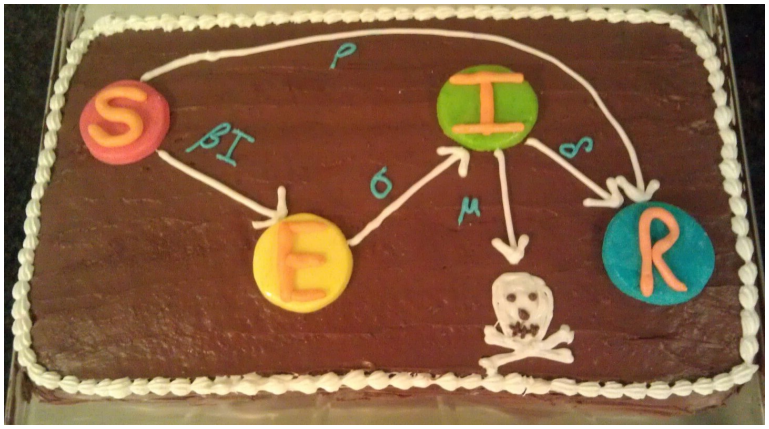
Difference-in-differences for infectious disease policy evaluation

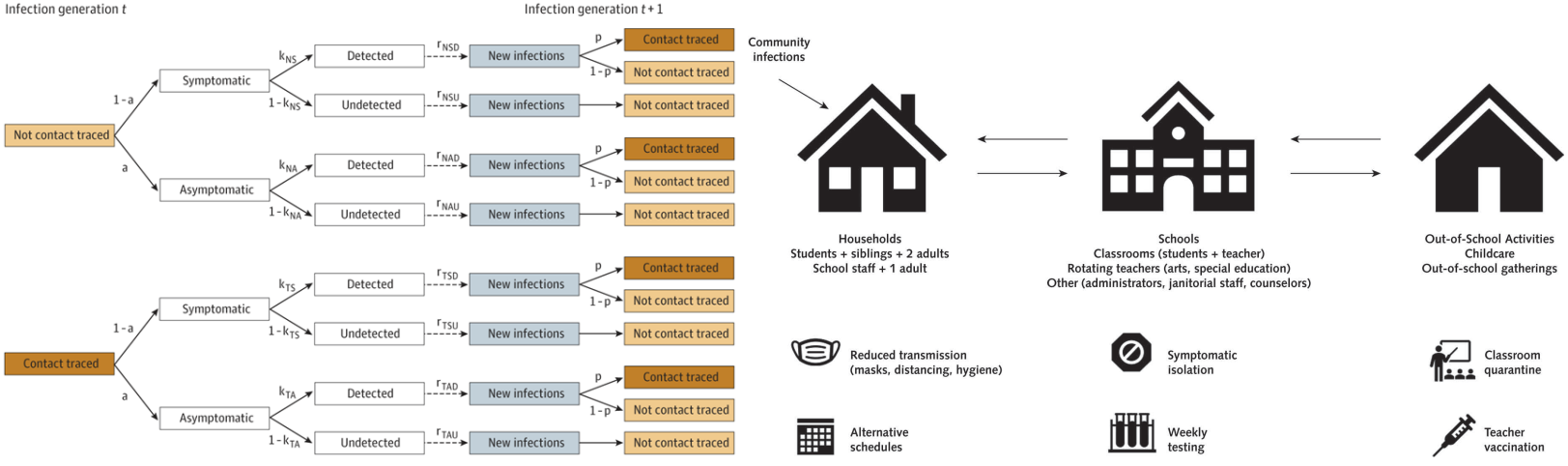
Alyssa Bilinski
Brown University

2025 Stata Biostatistics and Epidemiology Virtual Symposium
February 20, 2025

Infectious disease models are powerful tools for prediction and policy.

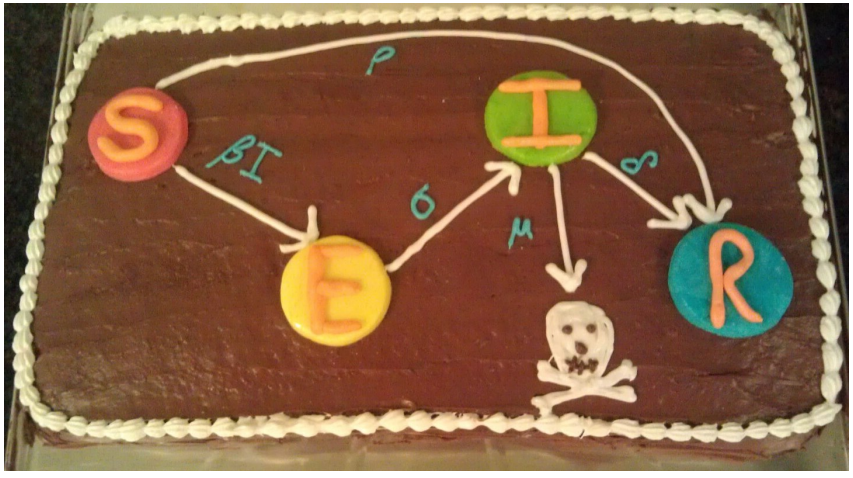
Infectious disease models are powerful tools for prediction and policy.





They have benefits...

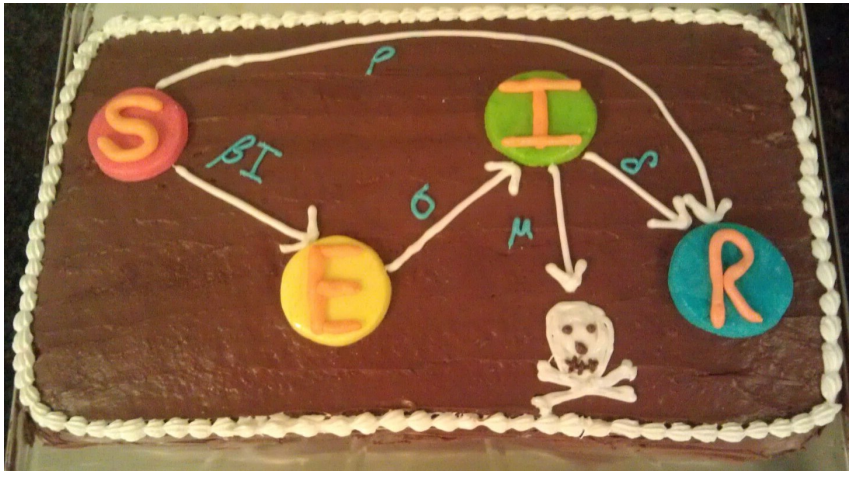
The Good



They have benefits...

The Good

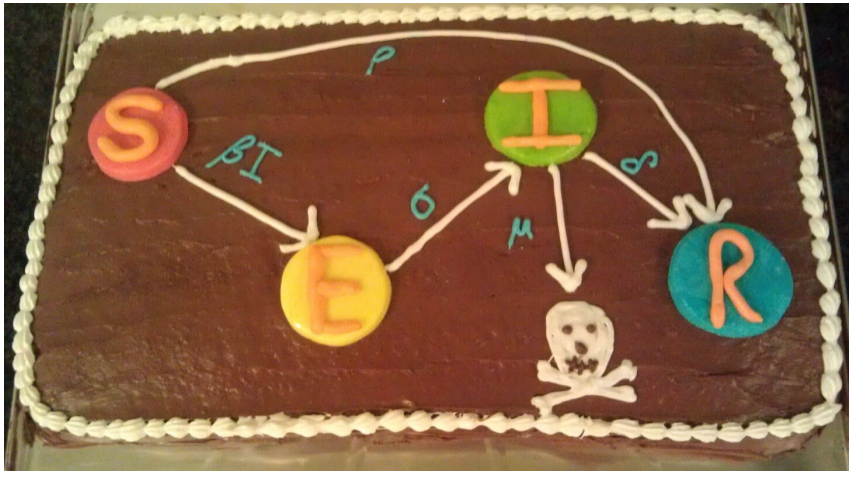
- Careful consideration of data-generating processes



They have benefits...

The Good

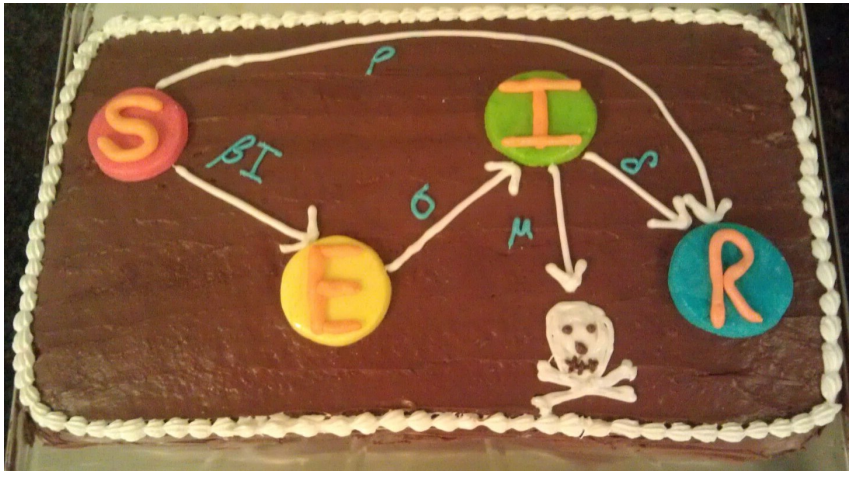
- Careful consideration of data-generating processes
- Make the most of limited/uncertain data



They have benefits...

The Good

- Careful consideration of data-generating processes
- Make the most of limited/uncertain data
- Forward-looking



They have benefits and drawbacks.

The Good

Careful consideration of data-generating processes

Make the most of limited/uncertain data

The Bad

They have benefits and drawbacks.

The Good

Careful consideration of data-generating processes

Make the most of limited/uncertain data

The Bad

- Typically scale effects from small or mechanistic studies

They have benefits and drawbacks.

The Good

Careful consideration of data-generating processes

Make the most of limited/uncertain data

The Bad

- Typically scale effects from small or mechanistic studies
- Often overly optimistic about the costs and effects of real-world programs

...and drawbacks.

The Good

Careful consideration of data-generating processes

Make the most of limited/uncertain data

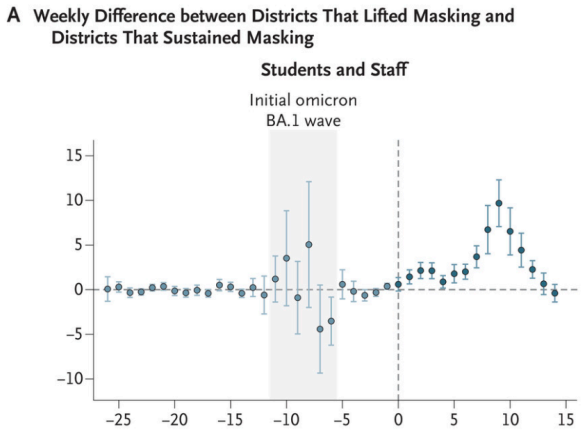
The Bad

- Typically scale effects from small or mechanistic studies
- Often overly optimistic about the costs and effects of real-world programs

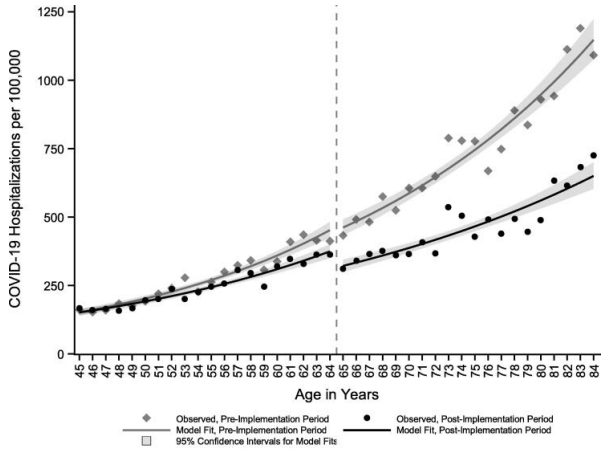
Salomon (2019):
We need "systematic reevaluation of the cost-effectiveness literature with reference to ***ex-post empirical evidence*** *on costs and effects in real-world programs."*

Observational causal inference methods can help.

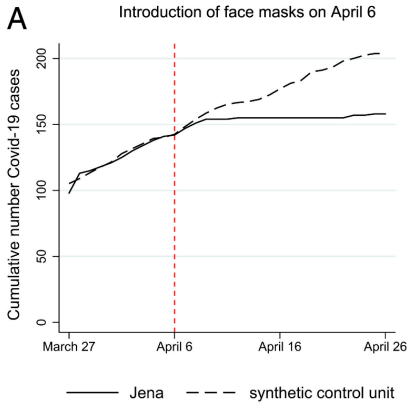
Difference-in-differences



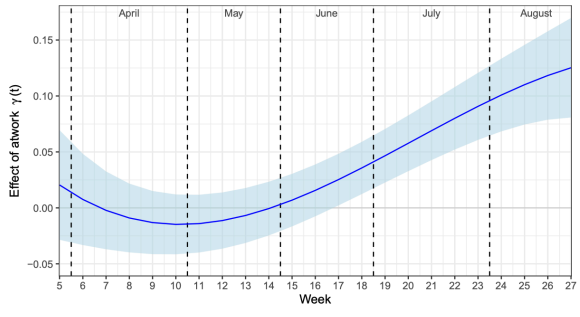
Regression discontinuity



Synthetic control methods



Instrumental variable methods



Observational causal inference methods can help.

Opinion | **What we need to know before we can end social distancing**

By Michael L. Barnett, Caroline O. Buckee and Yonatan H. Grad
April 1, 2020 at 10:23 a.m. EDT

But one unexpected benefit to the haphazard, staggered way that social distancing rolled out across the United States is that it created a host of natural experiments.

Observational causal inference methods can help.

Opinion | **What we need to know before we can end social distancing**

By Michael L. Barnett, Caroline O. Buckee and Yonatan H. Grad
April 1, 2020 at 10:23 a.m. EDT

But one unexpected benefit to the haphazard, staggered way that social distancing rolled out across the United States is that it created a host of natural experiments.

The Good

- Leverage real-world "exogenous" variation

Observational causal inference methods can help.

Opinion | **What we need to know before we can end social distancing**

By Michael L. Barnett, Caroline O. Buckee and Yonatan H. Grad
April 1, 2020 at 10:23 a.m. EDT

But one unexpected benefit to the haphazard, staggered way that social distancing rolled out across the United States is that it created a host of natural experiments.

The Good

- Leverage real-world "exogenous" variation
- Empirical counterfactuals from untreated units

They have their own drawbacks.

Journal of Econometrics
Volume 235, Issue 2, August 2023, Pages 2218-2244

What's trending in difference-in-differences? A synthesis of the recent econometrics literature ☆

The Good

Leverage real-world "exogenous" variation

Empirical counterfactuals from untreated units

The Bad

- Usually assume linear data-generating processes

They have their own drawbacks.

Journal of Econometrics
Volume 235, Issue 2, August 2023, Pages 2218-2244

What's trending in difference-in-differences? A synthesis of the recent econometrics literature ☆

The Good

Leverage real-world "exogenous" variation

Empirical counterfactuals from untreated units

The Bad

- Usually assume linear data-generating processes
- More willing to cop out given uncertainty

My work: Take the best of both worlds.

Infectious disease models

The Good

- Careful consideration of data-generating processes
- Make the most of limited/uncertain data

The Bad

Typically scale effects from small or mechanistic studies

Often overly optimistic about the costs and effects

Observational causal inference

The Good

- Leverage real-world "exogenous" variation
- Empirical counterfactuals from untreated units

The Bad

Usually assume linear data-generating processes

More willing to cop out given uncertainty

My work: Take the best of both worlds.

Infectious disease models

The Good

- Careful consideration of data-generating processes
- Make the most of limited/uncertain data

Observational causal inference

The Good

- Leverage real-world "exogenous" variation
- Empirical counterfactuals from untreated units

We develop **comprehensive theoretical architecture** for conducting observational policy evaluation and transporting results to inform projections. We also support its implementation in **public health practice**.

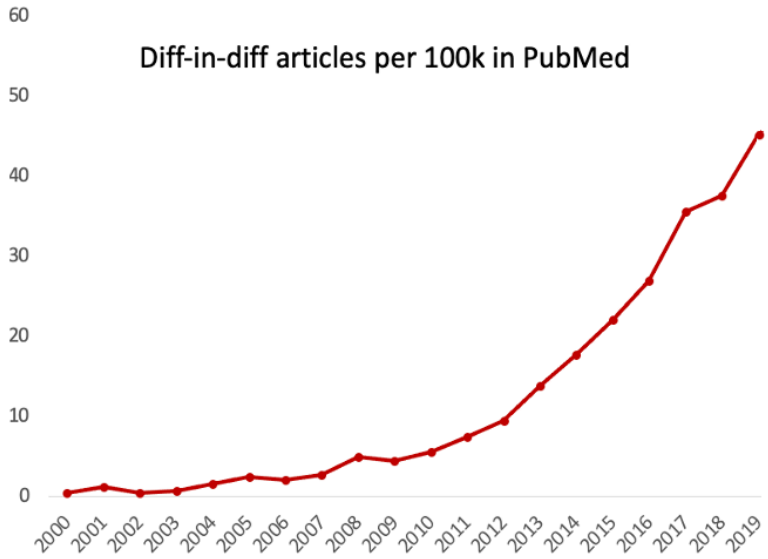
Today

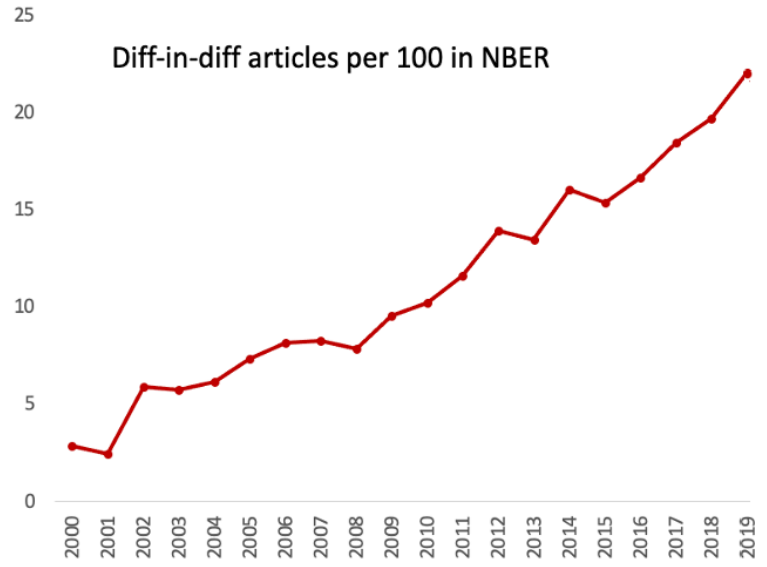
Difference-in-differences (DiD) for infectious disease outcomes

Shuo Feng
PhD Candidate, Biostatistics

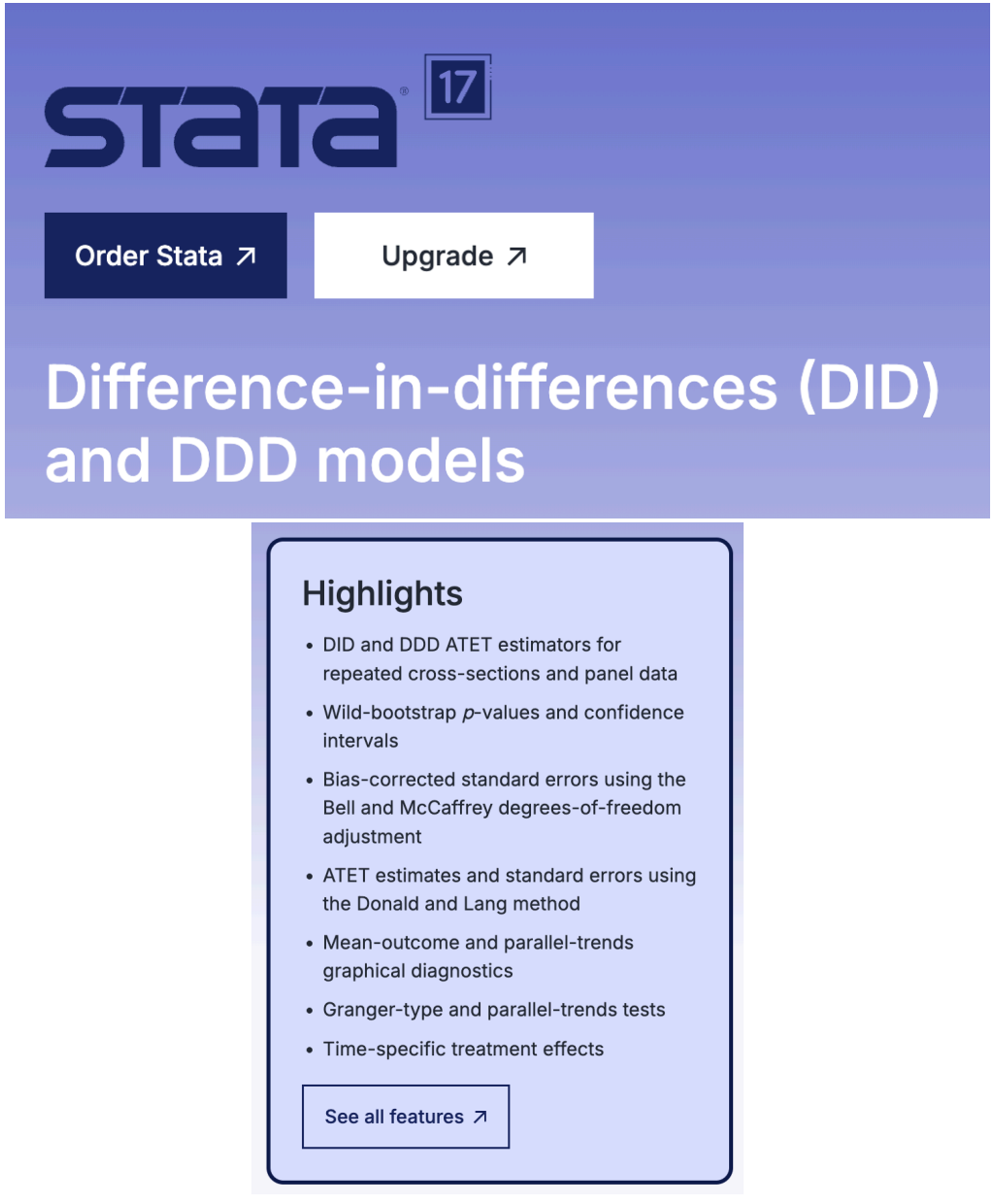
Supported in part by the Centers for Disease Control and Prevention though the Council of State and Territorial Epidemiologists (NU38OT000297-02)

DiD is popular.

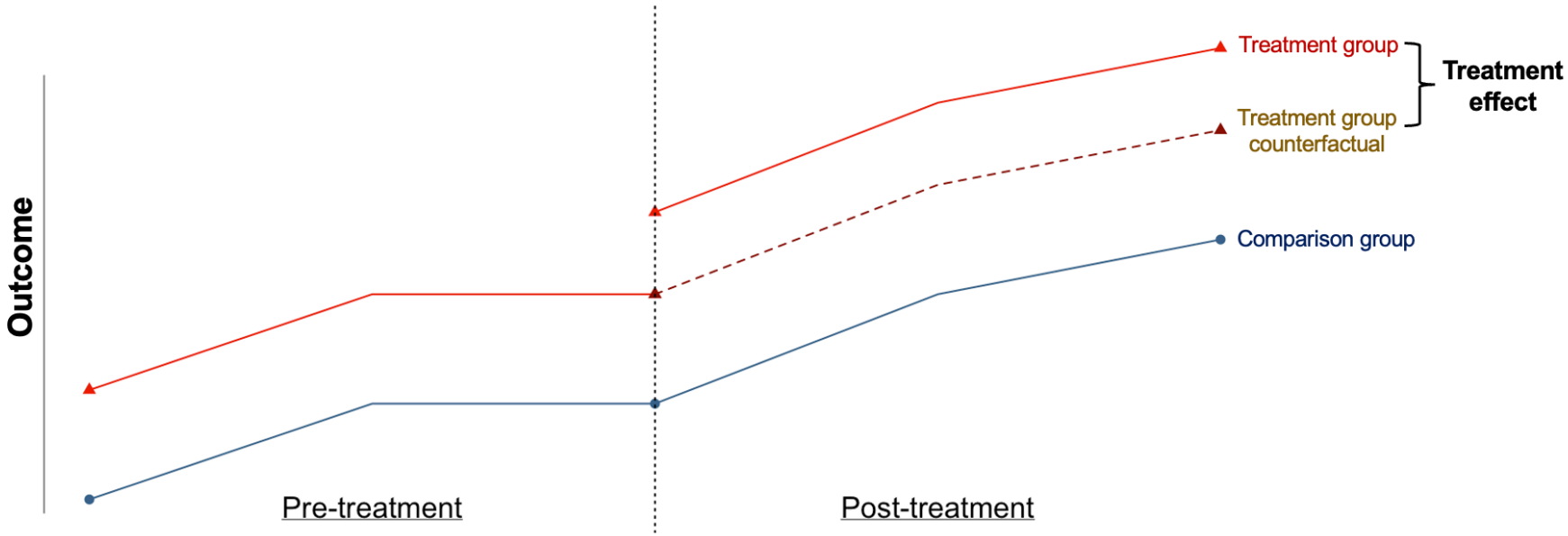




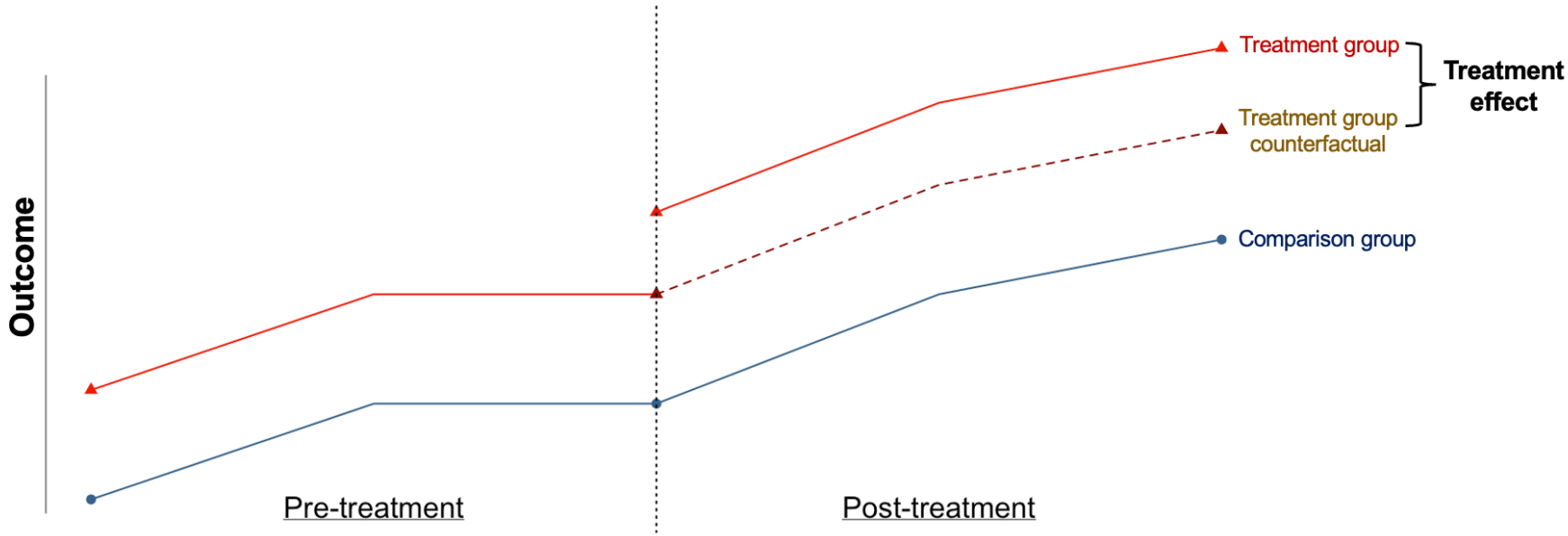
Especially among Stata users!



About DiD


Outcome
Pre-treatment
Post-treatment
Treatment group
Treatment group counterfactual
Comparison group
Treatment effect

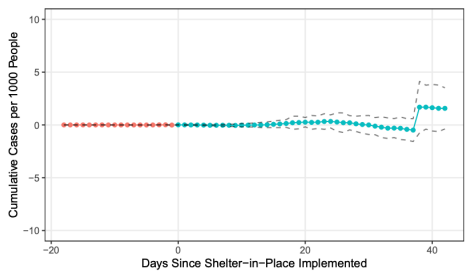
About DiD



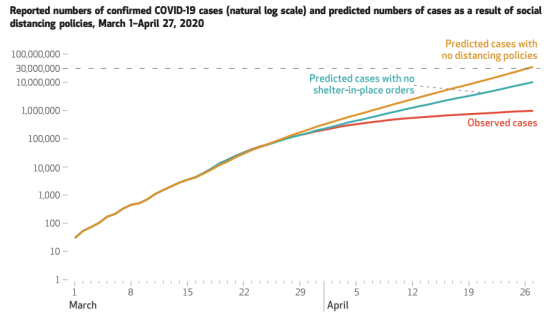
Parallel trends assumption (PTA): Treatment and comparison units were moving in parallel pre-intervention, and would have continued to do so absent the intervention (untestable).

DiD became even more popular during COVID-19.

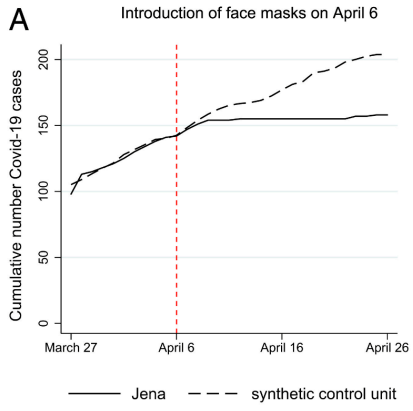
Stay-at-home orders
(Callaway & Li, 2023)



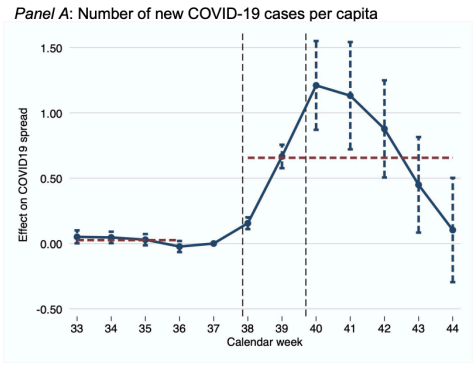
Social distancing
(Courtemanche et. al, 2020)



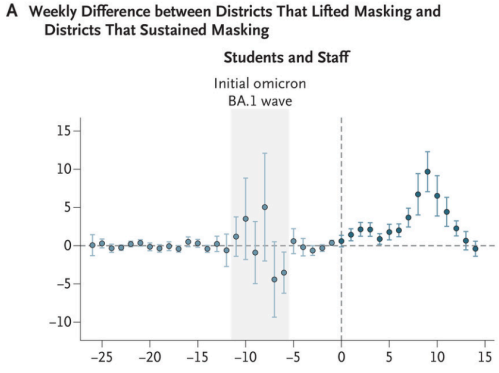
Mask mandates
(Mitze et. al, 2020)



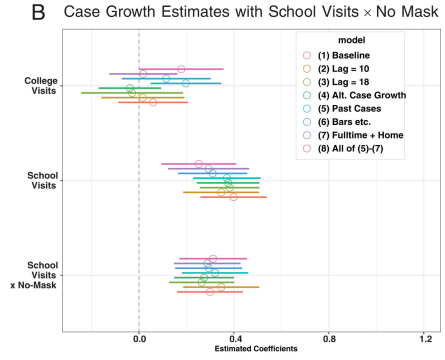
Contact tracing
(Fetzer & Graeber, 2021)



School mask mandates
(Cowger, et. al 2022)



School reopenings
(Chernozhukov, et. al. 2021)



COVID-19 policy evaluation

There was a lot of disagreement on how to use DiD to evaluate COVID-19 incidence and mortality:

- incidence
  - with matching
  - with synthetic controls
- log(incidence)
- log(incidence growth rate)
- log/log models

Do infectious disease dynamics mess up DiD?

Do infectious disease dynamics mess up DiD?

Why do we trust that trends are parallel, even if levels may differ?

Do infectious disease dynamics mess up DiD?

Why do we trust that trends are parallel, even if levels may differ?

- DiD is appealing because it allows researchers to estimate treatment effects, even absent comparison groups that exactly match the treatment group.
- But it's hard to know what the PTA confers about underlying data-generating processes.

Do infectious disease dynamics mess up DiD?

Why do we trust that trends are parallel, even if levels may differ?

- DiD is appealing because it allows researchers to estimate treatment effects, even absent comparison groups that exactly match the treatment group.
- But it's hard to know what the PTA confers about underlying data-generating processes.
- Infectious disease is a context with sufficiently deep theory to precisely interpret the PTA.

Objectives

What must we assume about treatment and comparison transmission dynamics for different DiD specifications to work well?

Objectives

What must we assume about treatment and comparison transmission dynamics for different DiD specifications to work well?

1. Formalize assumptions of different specifications from an epidemiological perspective

Objectives

What must we assume about treatment and comparison transmission dynamics for different DiD specifications to work well?

1. Formalize assumptions of different specifications from an epidemiological perspective
2. Propose robust alternatives guided by a structural framework

Objectives

What must we assume about treatment and comparison transmission dynamics for different DiD specifications to work well?

1. Formalize assumptions of different specifications from an epidemiological perspective
2. Propose robust alternatives guided by a structural framework
3. Characterize power trade-offs

Objectives

What must we assume about treatment and comparison transmission dynamics for different DiD specifications to work well?

1. Formalize assumptions of different specifications from an epidemiological perspective
2. Propose robust alternatives guided by a structural framework
3. Characterize power trade-offs
4. Re-analyze published examples

DiD Background

DiD specifications

Power

Examples

Discussion

Approach

1. Define an SIR transmission dynamic data-generating process for treatment and comparison units.

Approach

1. Define an SIR transmission dynamic data-generating process for treatment and comparison units.
2. Set up two-period, two-unit DiD model.

Approach

1. Define an SIR transmission dynamic data-generating process for treatment and comparison units.
2. Set up two-period, two-unit DiD model.
3. Formally derive mathematical conditions required for the PTA to hold in closed-form for different model specifications.

Approach

1. Define an SIR transmission dynamic data-generating process for treatment and comparison units.
2. Set up two-period, two-unit DiD model.
3. Formally derive mathematical conditions required for the PTA to hold in closed-form for different model specifications.
4. Interpret these in practical terms.

SIR model

We work with a stochastic SIR model. Assuming initial conditions S_0, I_0, R_0, for $t > 0$:

$$
\begin{aligned}
I^*_{t+1} &\sim Pois\left(\mu_t = \beta_t S_t \frac{I_t}{N}\right) \\
S_{t+1} &= S_t - I^*_{t+1} \\
I_{t+1} &= (1-\gamma) I_t + I^*_{t+1} \\
R_{t+1} &= R_t + \gamma I_t
\end{aligned}
$$

SIR model

We work with a stochastic SIR model. Assuming initial conditions S_0, I_0, R_0, for $t > 0$:

$$
\begin{aligned}
I^*_{t+1} &\sim Pois\left(\mu_t = \beta_t S_t \frac{I_t}{N}\right) \\
S_{t+1} &= S_t - I^*_{t+1} \\
I_{t+1} &= (1-\gamma) I_t + I^*_{t+1} \\
R_{t+1} &= R_t + \gamma I_t
\end{aligned}
$$

We assume I^*_{t+1} is Poisson distributed, but most results hold for any distribution with mean μ_t.

SIR assumptions

1. Closed, stable population of N individuals
2. Homogeneous mixing and transmission: $\beta_t \frac{I_t}{N}$ constant across individuals

Difference-in-differences

Assume a canonical DiD setup with two units, $D = \{0, 1\}$, and two time periods, $T = \{t_1, t_2\}$, with unit 1 treated at time t_2. Let $Y_{d,t}$ be the outcome of unit d at time t.

$$ATT = \mathbb{E}\left[Y_{1,t_2}(1) - Y_{1,t_2}(0)\right]$$

Difference-in-differences

Assume a canonical DiD setup with two units, $D = \{0, 1\}$, and two time periods, $T = \{t_1, t_2\}$, with unit 1 treated at time t_2. Let $Y_{d,t}$ be the outcome of unit d at time t.

$$ATT = \mathbb{E}\left[Y_{1,t_2}(1) - Y_{1,t_2}(0)\right]$$

Difference-in-differences

Assume a canonical DiD setup with two units, $D = \{0, 1\}$, and two time periods, $T = \{t_1, t_2\}$, with unit 1 treated at time t_2. Let $Y_{d,t}$ be the outcome of unit d at time t.

$$ATT = \mathbb{E}\left[Y_{1,t_2}(1) - Y_{1,t_2}(0)\right]$$

Difference-in-differences

Assume a canonical DiD setup with two units, $D = \{0, 1\}$, and two time periods, $T = \{t_1, t_2\}$, with unit 1 treated at time t_2. Let $Y_{d,t}$ be the outcome of unit d at time t.

$$ATT = \mathbb{E}\left[Y_{1,t_2}(1) - Y_{1,t_2}(0)\right]$$

To estimate this, we define the parallel trends assumption:

$$g\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - g\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right)$$
$$=$$
$$g\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - g\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right),$$

where $g(.)$ is continuous and monotonic.

Difference-in-differences

Assume a canonical DiD setup with two units, $D = \{0, 1\}$, and two time periods, $T = \{t_1, t_2\}$, with unit 1 treated at time t_2. Let $Y_{d,t}$ be the outcome of unit d at time t.

$$ATT = \mathbb{E}\left[Y_{1,t_2}(1) - Y_{1,t_2}(0)\right]$$

To estimate this, we define the parallel trends assumption:

$$g\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - g\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right)$$
$$=$$
$$g\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - g\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right),$$

where $g(.)$ is continuous and monotonic.

Model specifications

The parallel trends assumption is usually sensitive to functional form (Roth & Sant'Anna (2023)).

Model specifications

The parallel trends assumption is usually sensitive to functional form (Roth & Sant'Anna (2023)).

Specification	**Definition**	**Frequency**[i]
Incidence	$I^*_{d,t}$	17/29 (59%)
Log incidence	$\log(I^*_{d,t})$	10/29 (34%)
Log growth	$\log\left(\frac{I^*_{d,t}}{I^*_{d,t}}\right)$	2/29 (7%)

[i] Literature Review: all COVID-19 DiD analyses published in *JAMA* network journals, *New England Journal of Medicine*, *PNAS*, *Nature Research* journals, *Lancet*, *Health Affairs*, and *Health Economics* from 2020-2022.

Specifications

What would we have to assume about transmission dynamics under an SIR data-generating process for the PTA to hold?

Specification	Outcome	Link	Assumptions: Treatment vs. comparison parameters		
			Susceptible population	Initial infections	Effective contact rates
		$g(.)$	$(S_{d,0})$	$\mathbb{E}\left[Y_{d,t_1}\right]$	$\left(\beta_{d,t_1-1}, \ldots, \beta_{d,t_2-1}\right)$
Incidence	$Y_{d,t} = I^*_{d,t}$	identity			
Log incidence	$Y_{d,t} = I^*_{d,t}$	$\log$			
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right]}{\mathbb{E}\left[I^*_{d,t-1}\right]}$	$\log$			

Specifications

What would we have to assume about transmission dynamics under an SIR data-generating process for the PTA to hold?

Specification	Outcome	Link	Assumptions: Treatment vs. comparison parameters		
			Susceptible population	Initial infections	Effective contact rates
		$g(.)$	$(S_{d,0})$	$\mathbb{E}\left[Y_{d,t_1}\right]$	$\left(\beta_{d,t_1-1}, \ldots, \beta_{d,t_2-1}\right)$
Incidence	$Y_{d,t} = I^*_{d,t}$	identity			
Log incidence	$Y_{d,t} = I^*_{d,t}$	log			
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right]}{\mathbb{E}\left[I^*_{d,t-1}\right]}$	log			
Log R_t	$Y_{d,t} = R_{d,t}$	log			
Log β_t	$Y_{d,t} = \beta_{d,t}$	log			

Specifications

What would we have to assume about transmission dynamics under an SIR data-generating process for the PTA to hold?

Specification	Outcome	Link	Assumptions: Treatment vs. comparison parameters		
			Susceptible population	Initial infections	Effective contact rates
		$g(.)$	$(S_{d,0})$	$\mathbb{E}\left[Y_{d,t_1}\right]$	$\left(\beta_{d,t_1-1}, \ldots, \beta_{d,t_2-1}\right)$
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	$=$ or $S_{d,0} \to \infty$	$=$	$=$
Log incidence	$Y_{d,t} = I^*_{d,t}$	$\log$	$S_{d,0} \to \infty$		$=$
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right)}{\mathbb{E}\left[I^*_{d,t-1}\right]}$	$\log$	$S_{d,0} \to \infty$		constant ratio + $\beta_{d,t} = \beta_d$ or $\gamma = 1$
Log R_t	$Y_{d,t} = R_{d,t}$	$\log$	$S_{d,0} \to \infty$		constant ratio
Log β_t	$Y_{d,t} = \beta_{d,t}$	$\log$			constant ratio

$S_{d,0} \to \infty$

Specifications

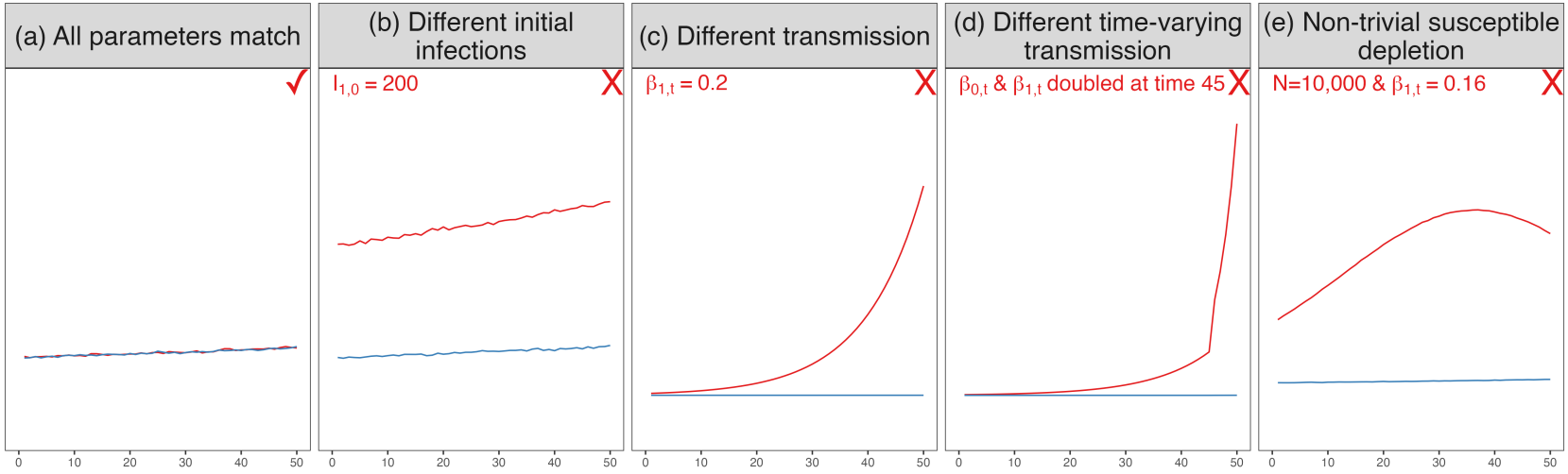
What would we have to assume about transmission dynamics under an SIR data-generating process for the PTA to hold?

Specification	Outcome	Link	Assumptions: Treatment vs. comparison parameters		
			Susceptible population	Initial infections	Effective contact rates
		$g(.)$	$(S_{d,0})$	$\mathbb{E}\left[Y_{d,t_1}\right]$	$\left(\beta_{d,t_1-1}, \ldots, \beta_{d,t_2-1}\right)$
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	$=$ or $S_{d,0} \to \infty$	$=$	$=$
Log incidence	$Y_{d,t} = I^*_{d,t}$	log	$S_{d,0} \to \infty$		$=$
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right)}{\mathbb{E}\left[I^*_{d,t-1}\right]}$	log	$S_{d,0} \to \infty$		constant ratio + $\beta_{d,t} = \beta_d$ or $\gamma = 1$
Log R_t	$Y_{d,t} = R_{d,t}$	log	$S_{d,0} \to \infty$		constant ratio
Log β_t	$Y_{d,t} = \beta_{d,t}$	log			constant ratio

$S_{d,0} \to \infty$

Incidence

$$g\left(\mathbb{E}\left[Y_{d,t}\right]\right) = \mathbb{E}\left(I^*_{d,t}\right)$$

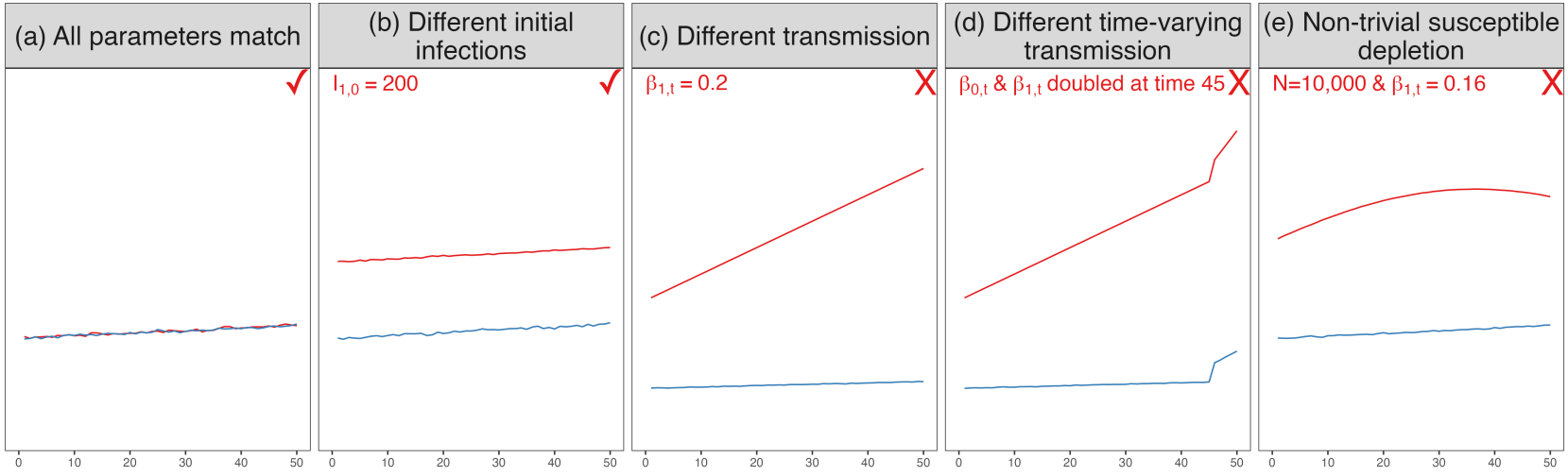


Base case: $I_0 = 50,\ \frac{S_0}{N} \approx 1,\ R_0 = \frac{\beta}{\gamma} = 1.6$

Formal conditions and derivation

Log incidence

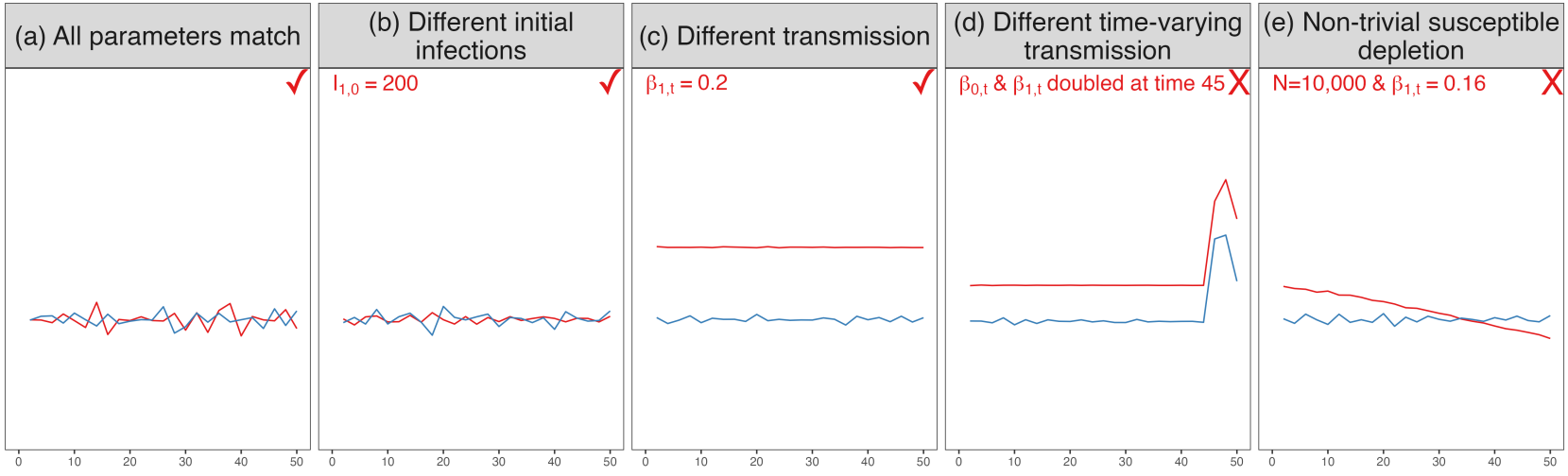
$$g\left(\mathbb{E}\left[Y_{d,t}\right]\right) = \log\left(\mathbb{E}\left[I^*_{d,t}\right]\right)$$



Formal conditions and derivation | Log transformations

Log growth

$$g\left(\mathbb{E}\left[Y_{d,t}\right]\right) = \log\left(\frac{\mathbb{E}\left[I^*_{j,t}\right]}{\mathbb{E}\left[I^*_{j,t-1}\right]}\right))$$



Formal conditions and derivation

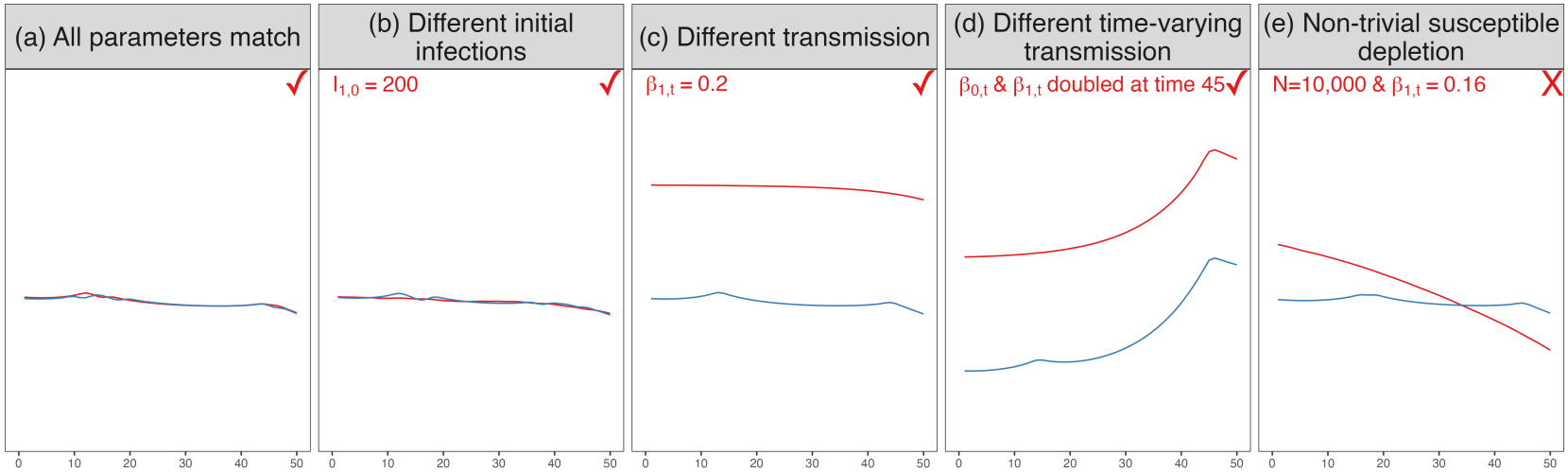
Log R_t

But...log growth approximates log R_t! Definition

Log R_t

But...log growth approximates log R_t! Definition

$$g\left(\mathbb{E}\left[Y_{d,t}\right]\right) = \log\left(\mathbb{E}\left[R_{d,t}\right]\right)$$



Formal conditions and derivation

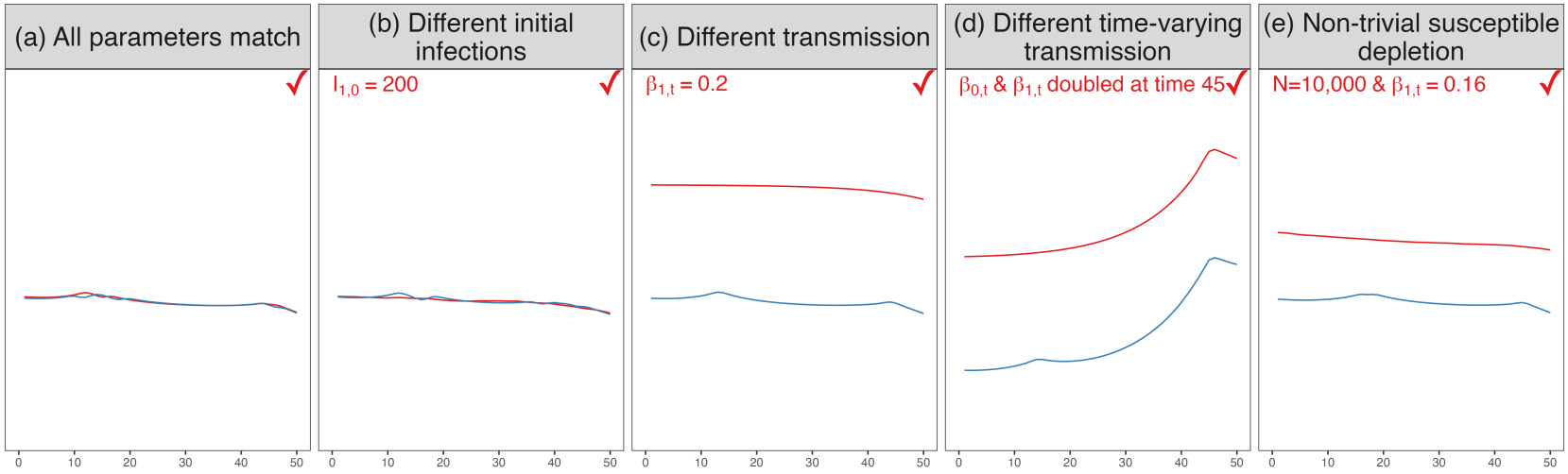
Log β_t

And then, why not model β_t directly?

Log β_t

And then, why not model β_t directly?

$$g\left(\mathbb{E}\left[Y_{d,t}\right]\right) = \log\left(\mathbb{E}\left[\beta_{d,t}\right]\right)$$



Formal conditions and derivation

ATT interpretations

Specification	**Outcome**	**Link** $g(.)$	**Interpretation of ATT**
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	Difference
Log incidence	$Y_{d,t} = I^*_{d,t}$	log	Percentage difference
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left(I^*_{d,t}\right)}{\mathbb{E}\left(I^*_{d,t-1}\right)}$	log	Percentage change
Log R_t	$Y_{d,t} = R_{d,t}$	log	Difference in avg. transmissions per infection
Log β_t	$Y_{d,t} = \beta_{d,t}$	log	Difference in effective contact rate

ATT interpretations

Specification	**Outcome**	**Link** $g(.)$	**Interpretation of ATT**
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	Difference
Log incidence	$Y_{d,t} = I^*_{d,t}$	log	Percentage difference
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left(I^*_{d,t}\right)}{\mathbb{E}\left(I^*_{d,t-1}\right)}$	log	Percentage change
Log R_t	$Y_{d,t} = R_{d,t}$	log	Difference in avg. transmissions per infection
Log β_t	$Y_{d,t} = \beta_{d,t}$	log	Difference in effective contact rate

$\longrightarrow$ We can also transform to average marginal effects.

Return

Inference

We conduct inference using the **wild score bootstrap** a generalization of the wild cluster bootstrap, which performs well with a **small number of clusters**.

Inference

We conduct inference using the **wild score bootstrap** a generalization of the wild cluster bootstrap, which performs well with a **small number of clusters**.

Key idea: In boostrap replicates, re-weight the score distribution based on an auxiliary cluster-level random variable with mean 0 and variance 1.

```
# run model
glm inc i.unit i.time 1.trt_post, family(poisson) link(log)

# wild cluster bootstrap
boottest 1.trt_post, cluster(unit) quietly
gen coef = _b[1.i.trt_post] in 1

# calculate p-value
gen p = r(p) in 1
keep coef p
```

Return

More complex models

I have never published an SIR model.

More complex models

I have never published an SIR model.

- Nearly all infectious disease models nest SIR models.
  - **SEIR models:** additional exposed state
  - **Agent-based models:** heterogeneity in individual agents

More complex models

I have never published an SIR model.

- Nearly all infectious disease models nest SIR models.
  - **SEIR models:** additional exposed state
  - **Agent-based models:** heterogeneity in individual agents
- In more complex models, there is no closed-form solution (requires exactly equality).

More complex models

I have never published an SIR model.

- Nearly all infectious disease models nest SIR models.
  - **SEIR models:** additional exposed state
  - **Agent-based models:** heterogeneity in individual agents
- In more complex models, there is no closed-form solution (requires exactly equality).
- But often SIR is a decent approximation (defaults to our propositions above).

Summary

Popular DiD specifications encode strong assumptions.

- Incidence requires identical expected outcome trajectories.
- Log incidence or log growth allows for different initial conditions under an "infinite susceptible population" assumption, but nevertheless requires strict conditions in transmission parameters.

Summary

Popular DiD specifications encode strong assumptions.

- Incidence requires identical expected outcome trajectories.
- Log incidence or log growth allows for different initial conditions under an "infinite susceptible population" assumption, but nevertheless requires strict conditions in transmission parameters.

R_t and β_t

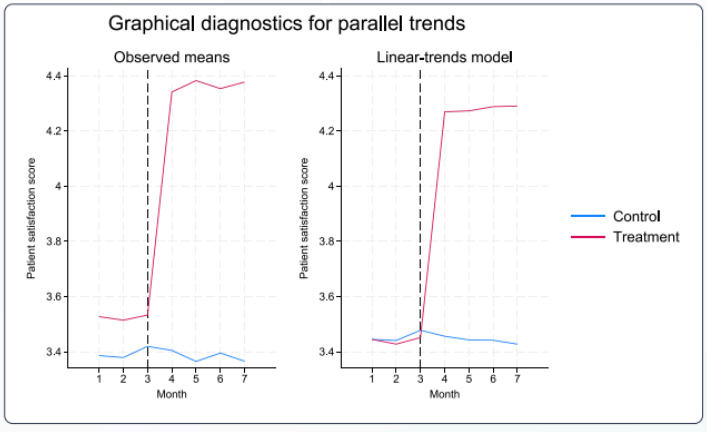
- Draw on fundamental epidemiological quantities
- More flexible than the established outcome specifications
- Can be estimated via MLE under SIR or with Wallinga-Teunis estimator under more complex transmission frameworks

ATT interpretations | Average marginal effects | Time step aggregation | Multiple units and time periods | R_t and β_t estimation | Inference

Summary

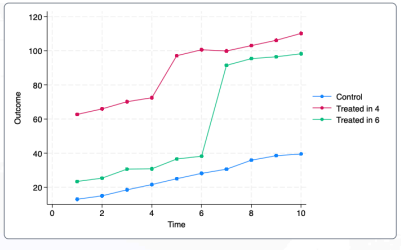
Graphical diagnostics

Our DID model assumes that the trends of **satis** for the control and treatment groups are parallel prior to the implementation of the new procedure. We can obtain a diagnostic of this assumption using **estat trendplot**.



Heterogeneous DID New

Heterogeneous DID estimates ATETs when treatment effects change over time and are different across cohorts. Use Stata's new **hdidregress** and **xthdidregress** commands to estimate ATETs for each cohort and time period with repeated cross-sectional data and panel data.



ATT interpretations | Average marginal effects | Time step aggregation

Multiple units and time periods | R_t and β_t estimation | Inference

DiD Background

DiD specifications

Power

Examples

Discussion

Power

What is the cost of using a more robust model?

Power

What is the cost of using a more robust model?

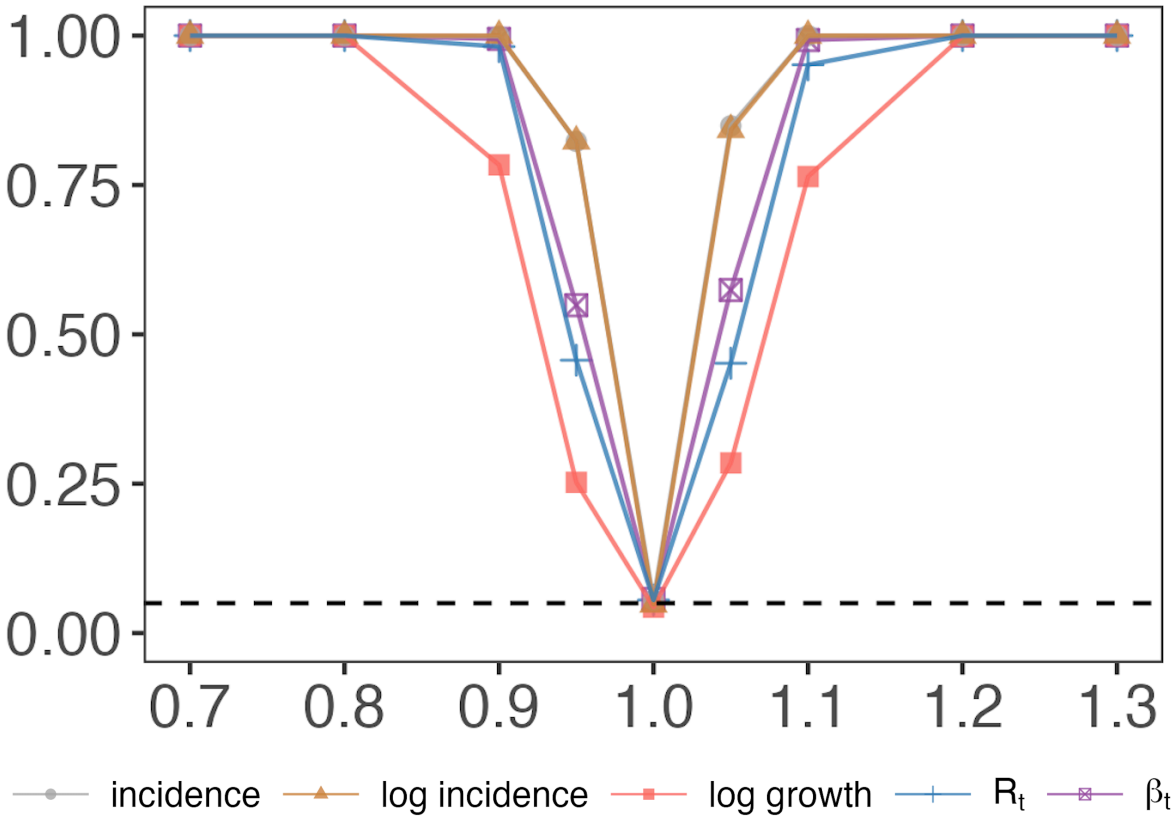
We ran simulations, generating data from an SIR model. Key input parameters:

- Ratio of β_t (the baseline effective contact rate) between treated and control groups: $\{1.0, 1.1\}$
- Effect size: 0.7-1.3 (a multiplicative factor on β_t)

We run each model and conduct inference with the wild score bootstrap, clustering at the unit level.

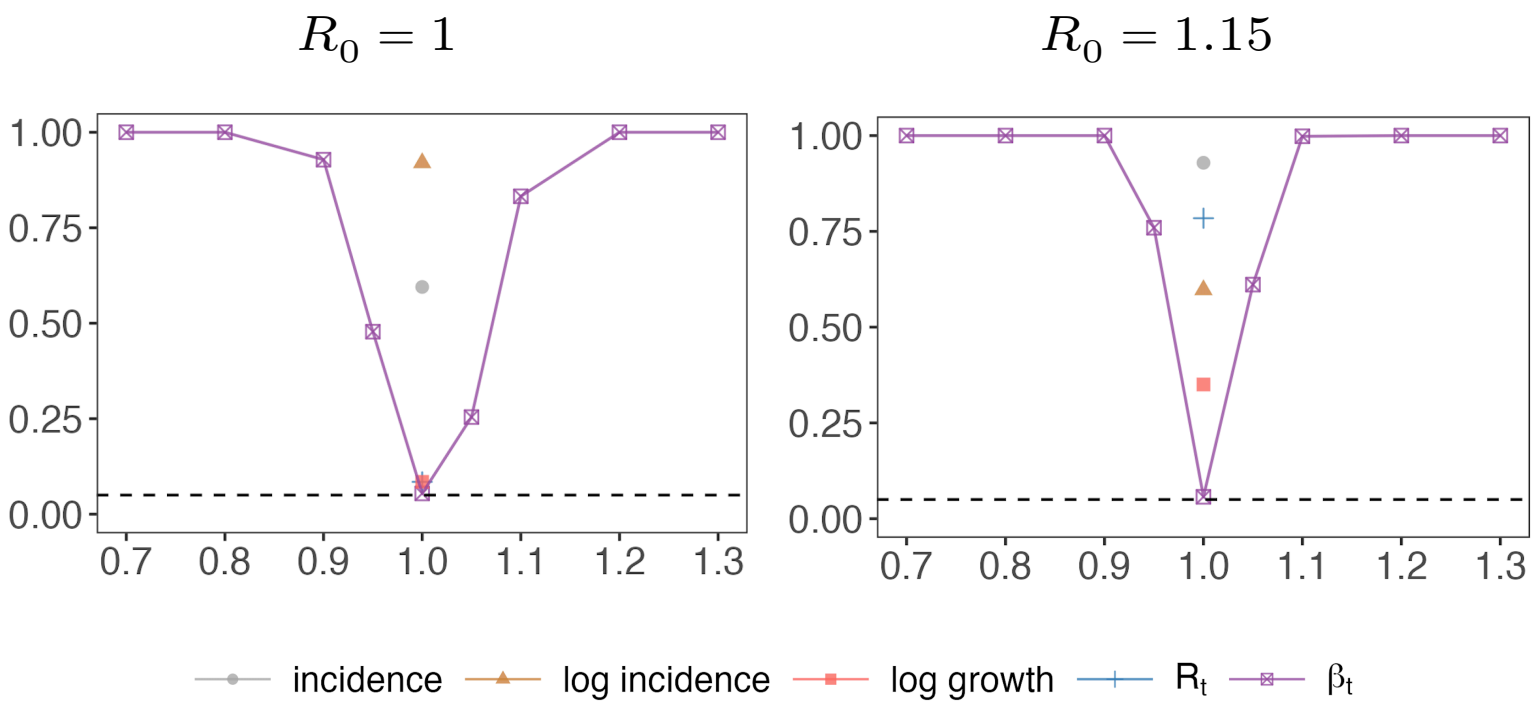
All parameters

Power (PTA holds for all outcome specifications)



All plots ($R_0 = 1$) All plots ($R_0 = 1.15$)

Power (β_t differs across treatment and comparison groups)



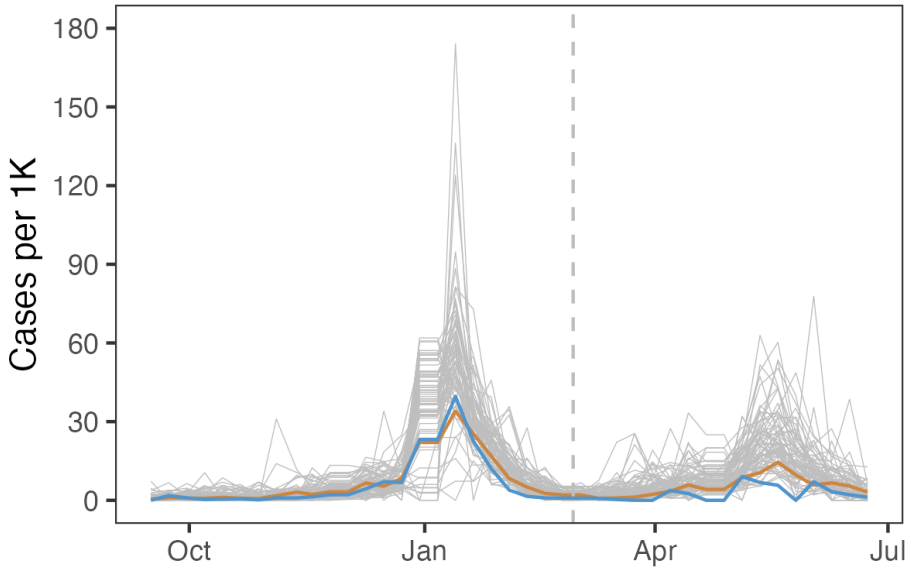
All plots ($R_0 = 1$) All plots ($R_0 = 1.15$)

Summary

1. **Use log by default:** Modeling log incidence does not substantially reduce power compared to modeling incidence.
2. **Log R_t and log β_t have greater power than log growth.**
3. **Beware susceptible depletion.** With non-trivial susceptible depletion, only log β_t could handle differences in effective contact rates.

School masking in Massachusetts

- On February 28, 2022, Massachusetts lifted the school state-level masking requirement.
- 2 districts did not lift mandates until June.



School masking in Massachusetts

Follow-up time	Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
15 weeks	Incidence	48.1^{*} $(38.9,\ 57.1)$	48.1^{*} $(38.9,\ 57.1)$
	Log incidence	1.19 $(0.90,\ 1.57)$	13.0 $(-31.3,\ 47.1)$
	Log growth	0.89 $(0.70,\ 1.13)$	-71.4 $(-3036.0,\ 120.5)$
5 weeks	Incidence	8.6^{*} $(5.7,\ 11.5)$	8.6^{*} $(5.7,\ 11.5)$
	Log incidence	1.62^{*} $(1.26,\ 2.09)$	6.9^{*} $(1.8,\ 10.8)$
	Log growth	0.97 $(0.77,\ 1.22)$	8.6 $(-10.9,\ 16.6)$

School masking in Massachusetts

Follow-up time	Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
15 weeks	Incidence	48.1^{*} $(38.9,\ 57.1)$	48.1^{*} $(38.9,\ 57.1)$
	Log incidence	1.19 $(0.90,\ 1.57)$	13.0 $(-31.3,\ 47.1)$
	Log growth	0.89 $(0.70,\ 1.13)$	-71.4 $(-3036.0,\ 120.5)$
5 weeks	Incidence	8.6^{*} $(5.7,\ 11.5)$	8.6^{*} $(5.7,\ 11.5)$
	Log incidence	1.62^{*} $(1.26,\ 2.09)$	6.9^{*} $(1.8,\ 10.8)$
	Log growth	0.97 $(0.77,\ 1.22)$	8.6 $(-10.9,\ 16.6)$

School masking in Massachusetts

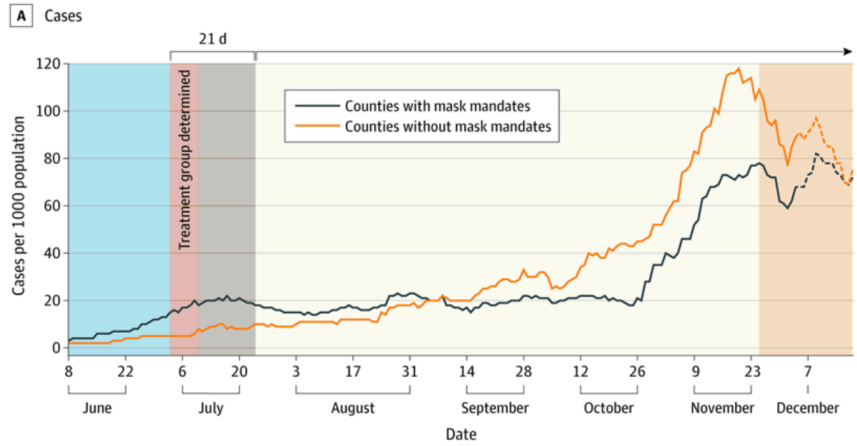
Follow-up time	Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
15 weeks	Incidence	48.1^{*} $(38.9,\ 57.1)$	48.1^{*} $(38.9,\ 57.1)$
	Log incidence	1.19 $(0.90,\ 1.57)$	13.0 $(-31.3,\ 47.1)$
	Log growth	0.89 $(0.70,\ 1.13)$	-71.4 $(-3036.0,\ 120.5)$
5 weeks	Incidence	8.6^{*} $(5.7,\ 11.5)$	8.6^{*} $(5.7,\ 11.5)$
	Log incidence	1.62^{*} $(1.26,\ 2.09)$	6.9^{*} $(1.8,\ 10.8)$
	Log growth	0.97 $(0.77,\ 1.22)$	8.6 $(-10.9,\ 16.6)$

School masking in Massachusetts

Follow-up time	Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
15 weeks	Incidence	48.1^{*} $(38.9,\ 57.1)$	48.1^{*} $(38.9,\ 57.1)$
	Log incidence	1.19 $(0.90,\ 1.57)$	13.0 $(-31.3,\ 47.1)$
	Log growth	0.89 $(0.70,\ 1.13)$	-71.4 $(-3036.0,\ 120.5)$
5 weeks	Incidence	8.6^{*} $(5.7,\ 11.5)$	8.6^{*} $(5.7,\ 11.5)$
	Log incidence	1.62^{*} $(1.26,\ 2.09)$	6.9^{*} $(1.8,\ 10.8)$
	Log growth	0.97 $(0.77,\ 1.22)$	8.6 $(-10.9,\ 16.6)$

Kansas mask mandates

- On July 3, 2020, Kansas passed an an executive order requiring masks.
- This was initially adopted only in 15 counties.



Kansas mask mandates

Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
Incidence	-21.6^{*} $(-28.3,\ -14.8)$	-21.6^{*} $(-28.3,\ -14.8)$
Log incidence	0.33^{*} $(0.22,\ 0.59)$	-55.1^{*} $(-101.2,\ -19.1)$
Log growth	0.96 $(0.74,\ 1.25)$	-9.5 $(-2139.8,\ 25.3)$
Log R_t	0.97 $(0.90,\ 1.05)$	-6.1 $(-22.2,\ 8.3)$
Log β_t	$0.95^{\dagger}$ $(0.86,\ 1.01)$	$-10.8^{\dagger}$ $(-36.0,\ 2.6)$

Kansas mask mandates

Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
Incidence	-21.6^{*} $(-28.3,\ -14.8)$	-21.6^{*} $(-28.3,\ -14.8)$
Log incidence	0.33^{*} $(0.22,\ 0.59)$	-55.1^{*} $(-101.2,\ -19.1)$
Log growth	0.96 $(0.74,\ 1.25)$	-9.5 $(-2139.8,\ 25.3)$
Log R_t	0.97 $(0.90,\ 1.05)$	-6.1 $(-22.2,\ 8.3)$
Log β_t	$0.95^{\dagger}$ $(0.86,\ 1.01)$	$-10.8^{\dagger}$ $(-36.0,\ 2.6)$

Kansas mask mandates

Outcome specification	ATT (95% CI)	Average marginal effect (95% CI)
Incidence	-21.6^{*} $(-28.3,\ -14.8)$	-21.6^{*} $(-28.3,\ -14.8)$
Log incidence	0.33^{*} $(0.22,\ 0.59)$	-55.1^{*} $(-101.2,\ -19.1)$
Log growth	0.96 $(0.74,\ 1.25)$	-9.5 $(-2139.8,\ 25.3)$
Log R_t	0.97 $(0.90,\ 1.05)$	-6.1 $(-22.2,\ 8.3)$
Log β_t	$0.95^{\dagger}$ $(0.86,\ 1.01)$	$-10.8^{\dagger}$ $(-36.0,\ 2.6)$

Contributions

1. Make explicit epidemiological assumptions embedded in popular DiD specifications with infectious disease outcomes
2. Propose robust specifications that can be generalized to more complex transmission dynamics
3. Characterize the bias-variance trade-off (e.g., logs as default)
4. Show that these differences are practically meaningful

How do we use observational estimates?

We estimate a 3-5% reduction in β_t from masking. How does this help us make future projections?

How do we use observational estimates?

We estimate a 3-5% reduction in β_t from masking. How does this help us make future projections?

1. Change benchmarks: 5-10% vs. 50-90%.

How do we use observational estimates?

We estimate a 3-5% reduction in β_t from masking. How does this help us make future projections?

1. Change benchmarks: 5-10% vs. 50-90%.
2. Improve model calibration: adherence? transmission location?

How do we use observational estimates?

We estimate a 3-5% reduction in β_t from masking. How does this help us make future projections?

1. Change benchmarks: 5-10% vs. 50-90%.
2. Improve model calibration: adherence? transmission location?
3. Guide data collection to inform context-specific updates.
4. Incorporate non-infectious disease outcomes.

How do we use observational estimates?

We estimate a 3-5% reduction in β_t from masking. How does this help us make future projections?

1. Change benchmarks: 5-10% vs. 50-90%.
2. Improve model calibration: adherence? transmission location?
3. Guide data collection to inform context-specific updates.
4. Incorporate non-infectious disease outcomes.

$\longrightarrow$ Cycles of projections and evaluation

Extensions to this work

Alternative causal inference methods

- Infectious disease dynamics are uniquely punishing!
- Synthetic control methods, regression discontinuity, spillovers

Extensions to this work

Alternative causal inference methods

- Infectious disease dynamics are uniquely punishing!
- Synthetic control methods, regression discontinuity, spillovers

Uncertainty

- If we require $p < 0.05$, it will be hard to act.
- Decision analytic methods for quantifying uncertainty
- Reduce researcher degrees of freedom (e.g., pre-analysis plans)

Extensions to this work

Alternative causal inference methods

- Infectious disease dynamics are uniquely punishing!
- Synthetic control methods, regression discontinuity, spillovers

Uncertainty

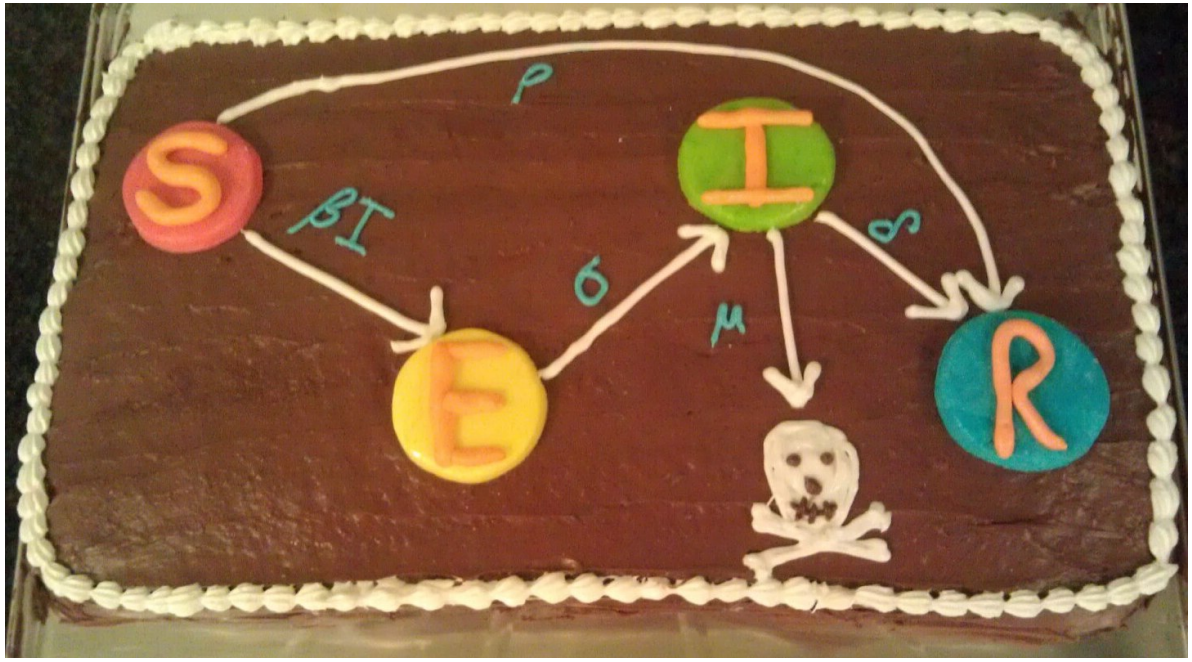
- If we require $p < 0.05$, it will be hard to act.
- Decision analytic methods for quantifying uncertainty
- Reduce researcher degrees of freedom (e.g., pre-analysis plans)

Applications

- CHAI: evaluation of malaria control efforts
- RIDOH: triggers for nursing home interventions
- NYC Health: framework for emergency policy evaluation

Thank you!

Feel free to reach out: alyssa_bilinski@brown.edu (especially if you are a Stata developer). Questions?



Expected incidence (SIR)

Proposition (Expected incidence)

Assuming an SIR data-generating process with initial conditions $\{S_{d,0}, I_{d,0}, R_{d,0}\}$, *expected incidence at times* $t+1$ *can be written:*

$$\mathbb{E}\left[I^*_{d,t+1}\right] = \frac{\beta_{d,t}}{N}\left(S_{d,0} + \frac{(1-\gamma)}{\beta_{d,t-1}}\right)\mathbb{E}\left[I^*_{d,t}\right] - \epsilon_t,$$

where $\epsilon_t = (1-\gamma)\mathbb{E}\left[I_{d,t-1}I^*_{d,t}\right] - \sum_{j=1}^{t}\mathbb{E}\left[I^*_{d,t}I^*_{d,j}\right]$.

This result suggests that even with t_1 and t_2 as adjacent time-steps, we cannot straightforwardly write $\mathbb{E}\left[I^*_{d,t+1}\right]$ as an additive or multiplicative function of $\mathbb{E}\left[I^*_{d,t+1}\right]$, implying a requirement of equivalent data-generating processes for the parallel trends assumption to hold with incidence or a log transformation.

$S_{d,0} \to \infty$

Expected incidence (SIR infinite population)

Proposition (Expected incidence (infinite susceptible population))

Assuming an SIR data-generating process, with initial conditions $\{S_0, I_0, R_0\}$*, for* $t \geq 1$*,*

$$
\begin{aligned}
E\left[\lim_{S_{d,0}\to\infty} I^*_{d,t+1}\right] &= \beta_{d,t} \prod_{k=0}^{t-1} \left(1 - \gamma + \beta_{d,k}\right) I_{d,0} \\
&= \lim_{S_{d,0}\to\infty} E\left[I^*_{d,t+1}\right]
\end{aligned}
$$

$S_{d,0} \to \infty$

Incidence: Exponential model

Recall $I^*_{j,g} = I_{j,0} r_j^g$. The parallel trends assumption holds when:

$$\begin{aligned}
\mathbb{E}\left[Y_{1,t_2}(0) - Y_{1,t_1}(0)\right] &= \mathbb{E}\left[Y_{0,t_2}(0) - Y_{0,t_1}(0)\right] \iff \\
I_{1,0}\left(r_1^{t_2} - r_1^{t_1}\right) &= I_{0,0}\left(r_0^{t_2} - r_0^{t_1}\right) \iff \\
Y_{1,t_1}\left({r_1}^{t_2-t_1} - 1\right) &= Y_{0,t_1}\left({r_0}^{t_2-t_1} - 1\right)
\end{aligned}$$

Incidence: Exponential model

Recall $I^*_{j,g} = I_{j,0} r_j^g$. The parallel trends assumption holds when:

$$\begin{aligned}
\mathbb{E}\left[Y_{1,t_2}(0) - Y_{1,t_1}(0)\right] &= \mathbb{E}\left[Y_{0,t_2}(0) - Y_{0,t_1}(0)\right] \iff \\
I_{1,0}\left(r_1^{t_2} - r_1^{t_1}\right) &= I_{0,0}\left(r_0^{t_2} - r_0^{t_1}\right) \iff \\
Y_{1,t_1}\left({r_1}^{t_2-t_1} - 1\right) &= Y_{0,t_1}\left({r_0}^{t_2-t_1} - 1\right)
\end{aligned}$$

Incidence: Exponential model

Recall $I^*_{j,g} = I_{j,0} r_j^g$. The parallel trends assumption holds when:

$$\begin{aligned}
\mathbb{E}\left[Y_{1,t_2}(0) - Y_{1,t_1}(0)\right] &= \mathbb{E}\left[Y_{0,t_2}(0) - Y_{0,t_1}(0)\right] \iff \\
I_{1,0}\left(r_1^{t_2} - r_1^{t_1}\right) &= I_{0,0}\left(r_0^{t_2} - r_0^{t_1}\right) \iff \\
Y_{1,t_1}\left({r_1}^{t_2-t_1} - 1\right) &= Y_{0,t_1}\left({r_0}^{t_2-t_1} - 1\right)
\end{aligned}$$

Incidence

Proposition (Parallel trends: Incidence)

Assuming an SIR data-generating process and an incidence model specification $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\mathbb{E}\left[Y_{1,t_2}(0)\right] - \mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty}\left(\mathbb{E}\left[Y_{0,t_2}(0)\right] - \mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$

$$\beta^*_{1,0,t_1} I_{1,0}\left(\beta^*_{1,t_1,t_2} - 1\right) = \beta^*_{0,0,t_1} I_{0,0}\left(\beta^*_{0,t_1,t_2} - 1\right),$$

$$\text{where } \beta^*_{d,0,t_1} = \beta_{d,t_1-1}\prod_{k=0}^{t_1-2}\left(1-\gamma+\beta_{d,k}\right),$$

$$\beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}}\prod_{k=t_1-1}^{t_2-2}\left(1-\gamma+\beta_{d,k}\right)$$

Return

Incidence

Proposition (Parallel trends: Incidence)

Assuming an SIR data-generating process and an incidence model specification $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\mathbb{E}\left[Y_{1,t_2}(0)\right] - \mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty}\left(\mathbb{E}\left[Y_{0,t_2}(0)\right] - \mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$

$$\beta^*_{1,0,t_1} I_{1,0}\left(\beta^*_{1,t_1,t_2} - 1\right) = \beta^*_{0,0,t_1} I_{0,0}\left(\beta^*_{0,t_1,t_2} - 1\right),$$

$$\text{where } \beta^*_{d,0,t_1} = \beta_{d,t_1-1}\prod_{k=0}^{t_1-2}\left(1-\gamma+\beta_{d,k}\right),$$

$$\beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}}\prod_{k=t_1-1}^{t_2-2}\left(1-\gamma+\beta_{d,k}\right)$$

Return

Incidence

Proposition (Parallel trends: Incidence)

Assuming an SIR data-generating process and an incidence model specification $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\mathbb{E}\left[Y_{1,t_2}(0)\right] - \mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty}\left(\mathbb{E}\left[Y_{0,t_2}(0)\right] - \mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$

$$\mathbb{E}\left[I^*_{1,t_1}\right]\left(\beta^*_{1,t_1,t_2} - 1\right) = \mathbb{E}\left[I^*_{0,t_1}\right]\left(\beta^*_{0,t_1,t_2} - 1\right),$$

$$\text{where } \beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}} \prod_{k=t_1-1}^{t_2-2}\left(1 - \gamma + \beta_{d,k}\right)$$

Return

Incidence

Proposition (Parallel trends: Incidence)

Assuming an SIR data-generating process and an incidence model specification $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\mathbb{E}\left[Y_{1,t_2}(0)\right] - \mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty}\left(\mathbb{E}\left[Y_{0,t_2}(0)\right] - \mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$

$$\mathbb{E}\left[I^*_{1,t_1}\right]\left(\beta^*_{1,t_1,t_2} - 1\right) = \mathbb{E}\left[I^*_{0,t_1}\right]\left(\beta^*_{0,t_1,t_2} - 1\right),$$

$$\text{where } \beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}} \prod_{k=t_1-1}^{t_2-2}\left(1 - \gamma + \beta_{d,k}\right)$$

Return

Proof.

For $t_1, t_2 \geq 2$ and $d \in \{0, 1\}$, we have:

$$\lim_{S_{d,0} \to \infty} \left(\mathbb{E}\left[Y_{d,t_2}(0) \right] - \mathbb{E}\left[Y_{d,t_1}(0) \right] \right)$$

Substituting from Proposition 2 $\longrightarrow$

$$= \beta_{d,t_2-1} \prod_{k=0}^{t_2-2} \left(1 - \gamma + \beta_{d,k}\right) I_{d,0} - \beta_{d,t_1-1} \prod_{k=0}^{t_1-2} \left(1 - \gamma + \beta_{d,k}\right) I_{d,0}$$

Rearranging terms $\longrightarrow$

$$= I_{d,0} \prod_{k=0}^{t_1-2} \left(1 - \gamma + \beta_{d,k}\right) \left(\beta_{d,t_2-1} \prod_{k=t_1-1}^{t_2-2} \left(1 - \gamma + \beta_{d,k}\right) - \beta_{d,t_1-1} \right)$$

Collecting terms $\longrightarrow$

$$= \beta^*_{d,0,t_1} I_{d,0} \left(\beta^*_{d,t_1,t_2} - 1 \right),$$

where $\beta^*_{d,0,t_1} = \beta_{d,t_1-1} \prod_{k=0}^{t_1-2} \left(1 - \gamma + \beta_{d,k}\right)$, $\beta^*_{d,t_1,t_2} = \dfrac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}} \prod_{k=t_1-1}^{t_2-2} \left(1 - \gamma + \beta_{d,k}\right)$

Substituting the above expression into the parallel trends condition, we obtain [a]:

$$LHS = \lim_{S_{1,0} \to \infty} \mathbb{E}\left[Y_{1,t_2}(0) - Y_{1,t_1}(0) \right] = \beta^*_{1,0,t_1} I_{1,0} \left(\beta^*_{1,t_1,t_2} - 1 \right), \text{ and}$$

$$RHS = \lim_{S_{0,0} \to \infty} \mathbb{E}\left[Y_{0,t_2}(0) - Y_{0,t_1}(0) \right] = \beta^*_{0,0,t_1} I_{0,0} \left(\beta^*_{0,t_1,t_2} - 1 \right)$$

□

Log incidence

Proposition (Parallel trends: Log incidence)

Assuming an SIR data-generating process and a log incidence model specification $\left(Y_{d,t} = I^*_{d,t},\ g(\cdot) = \log(\cdot)\right)$, *the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty} \log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty} \log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right)$$

$$\iff$$

$$\beta^*_{1,t_1,t_2} = \beta^*_{0,t_1,t_2},$$

$$\text{where } \beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}} \prod_{k=t_1-1}^{t_2-2} \left(1 - \gamma + \beta_{d,k}\right)$$

Return

Proof.

For $t_1, t_2 \geq 2$ and $d \in \{0, 1\}$, we expand as follows:

$$\lim_{S_{d,0} \to \infty} \log\left(\mathbb{E}\left[Y_{d,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{d,t_1}(0)\right]\right)$$

Substituting in from Proposition 2 $\longrightarrow$

$$= \log\left(\beta_{d,t_2-1} \prod_{k=0}^{t_2-2} (1 - \gamma + \beta_{d,k})\, I_{d,0}\right) - \log\left(\beta_{d,t_1-1} \prod_{k=0}^{t_1-2} (1 - \gamma + \beta_{d,k})\, I_{d,0}\right)$$

Dividing out common terms $\longrightarrow$

$$= \log\left(\frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}} \prod_{k=t_1-1}^{t_2-2} (1 - \gamma + \beta_{d,k})\right)$$

$$= \log\left(\beta^*_{d,t_1,t_2}\right)$$

Therefore, the "infinite susceptible population" parallel trends assumption (Eq. **??**) holds if and only if

$$\lim_{S_{1,0} \to \infty} \log\left(\mathbb{E}\left[Y_{1,t_2}\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}\right]\right) = \lim_{S_{0,0} \to \infty} \log\left(\mathbb{E}\left[Y_{0,t_2}\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}\right]\right) \iff$$

$$\log\left(\beta^*_{1,t_1,t_2}\right) = \log\left(\beta^*_{0,t_1,t_2}\right) \iff$$

$$\beta^*_{1,t_1,t_2} = \beta^*_{0,t_1,t_2}$$

□

Log transformations

We can approximate $\mathbb{E}\left[\log(I^*_{d,t})\right]$ with a second-order Taylor series expansion:

$$\begin{aligned}\mathbb{E}\left[\log(I^*_{d,t})\right] &\approx \log\left(\mathbb{E}(I^*_t)\right) - \frac{Var(I^*_{d,t})}{2\mathbb{E}(I_{d,t})^{*2}} \\ &\approx \log\left(\mathbb{E}(I^*_t)\right) - \frac{1}{2\mathbb{E}(I_{d,t})}\end{aligned}$$

Return

Parallel trends: Log incidence

Proposition (Parallel trends: Log growth)

Assuming an SIR data-generating process and a log growth model specification $\left(Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right]}{\mathbb{E}\left[I^*_{d,t-1}\right]},\ g(\cdot) = \log(\cdot)\right)$*, the "infinite susceptible population" parallel trends assumption holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right)\right) =$$

$$\lim_{S_{0,0}\to\infty}\left(\log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right)\right) \iff$$

$$\log\left(\frac{\beta_{1,t_2-1}}{\beta_{1,t_1-1}}\right) - \log\left(\frac{\beta_{1,t_2-2}}{\beta_{1,t_1-2}}\right) + \log\left(\frac{1-\gamma+\beta_{1,t_2-2}}{1-\gamma+\beta_{1,t_1-2}}\right) =$$

$$\log\left(\frac{\beta_{0,t_2-1}}{\beta_{0,t_1-1}}\right) - \log\left(\frac{\beta_{0,t_2-2}}{\beta_{0,t_1-2}}\right) + \log\left(\frac{1-\gamma+\beta_{0,t_2-2}}{1-\gamma+\beta_{0,t_1-2}}\right)$$

Return

Proof.

For $t_1, t_2 \geq 2$ and $d \in \{0, 1\}$, we have:

$$\lim_{S_{d,0} \to \infty} \log\left(\mathbb{E}\left[Y_{d,t_2}\right]\right)$$

Substituting in from Proposition 2 $\longrightarrow$

$$= \log\left(\beta_{d,t_2-1} \prod_{k=1}^{t_2-2} \left(1 - \gamma + \beta_{d,k}\right) I_{d,0}\right) - \log\left(\beta_{d,t_2-2} \prod_{k=1}^{t_2-3} \left(1 - \gamma + \beta_{d,k}\right) I_{d,0}\right)$$

Simplifying $\longrightarrow$

$$= \log\left(\frac{\beta_{d,t_2-1}}{\beta_{d,t_2-2}} \left(1 - \gamma + \beta_{d,t_2-2}\right)\right)$$

Similarly, $\log\left(\mathbb{E}\left[Y_{d,t_1}\right]\right) = \log\left(\frac{\beta_{d,t_1-1}}{\beta_{d,t_1-2}} \left(1 - \gamma + \beta_{d,t_1-2}\right)\right)$.

Therefore,

$$\lim_{S_{d,0} \to \infty} \log\left(\mathbb{E}\left[Y_{d,t_2}\right]\right) - \log\left(\mathbb{E}\left[Y_{d,t_1}\right]\right)$$

Substituting from above $\longrightarrow$

$$= \log\left(\frac{\beta_{d,t_2-1}}{\beta_{d,t_2-2}} \left(1 - \gamma + \beta_{d,t_2-2}\right)\right) - \log\left(\frac{\beta_{d,t_1-1}}{\beta_{d,t_1-2}} \left(1 - \gamma + \beta_{d,t_1-2}\right)\right)$$

Rearranging terms $\longrightarrow$

$$= \log\left(\frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}}\right) - \log\left(\frac{\beta_{d,t_2-2}}{\beta_{d,t_1-2}}\right) + \log\left(\frac{1 - \gamma + \beta_{d,t_2-2}}{1 - \gamma + \beta_{d,t_1-2}}\right)$$

Substituting the above equation back to both sides of the parallel trends assumption completes the proof. □

Definition of R_t

Proposition (Cohort definition of R_t)

Assume that the effective reproduction number is measured over a generation interval of length $\frac{1}{\gamma}$ for the cohort I_t^* becoming infectious at time t. We define the cohort effective reproduction number:

$$R_{d,t} = \sum_{j=t}^{\infty} (1-\gamma)^{j-t} \, \beta_{d,j} \, \frac{S_{d,j}}{N}$$

.

Return

Definition of R_t

Proposition (Cohort definition of R_t)

Assume that the effective reproduction number is measured over a generation interval of length $\frac{1}{\gamma}$ for the cohort I_t^* becoming infectious at time t. We define the cohort effective reproduction number:

$$R_{d,t} = \sum_{j=t}^{\infty} (1-\gamma)^{j-t} \, \beta_{d,j} \, \frac{S_{d,j}}{N}$$

.

Return

Parallel trends: Log R_t

Proposition (Parallel trends: Log R_t)

Assuming an SIR data-generating process, log-transformed effective reproduction number model specification $\left(Y_{d,t} = \log\left(R_{d,t}\right), g(\cdot) = \log(\cdot)\right)$*, the "infinite susceptible population" parallel trends assumption holds for all* $t_1, t_2 > t-1$ *if and only if*

$$\lim_{S_{1,0}\to\infty} \log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \lim_{S_{0,0}\to\infty} \log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$

$$\log\left(\frac{\sum_{j=t_2}^{\infty}(1-\gamma)^{j-t_2}\,\beta_{1,j}}{\sum_{j=t_1}^{\infty}(1-\gamma)^{j-t_1}\,\beta_{1,j}}\right) = \log\left(\frac{\sum_{j=t_2}^{\infty}(1-\gamma)^{j-t_2}\,\beta_{0,j}}{\sum_{j=t_1}^{\infty}(1-\gamma)^{j-t_1}\,\beta_{0,j}}\right)$$

Return

Proof.

In Proposition 9, we defined $R_t = R_{d,t} = \sum_{j=t}^{\infty} (1-\gamma)^{j-t}\, \beta_{d,j}\, \frac{S_{d,j}}{N}$. Substituting this formulation of R_t into the parallel trends assumption, we obtain:

$$\lim_{S_{1,0}\to\infty} \log\left(\mathbb{E}\left[Y_{1,t+1}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t}(0)\right]\right) = \lim_{S_{0,0}\to\infty} \log\left(\mathbb{E}\left[Y_{0,t+1}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t}(0)\right]\right) \iff$$

$$\lim_{S_{1,0}\to\infty} \log\left(R_{1,t+1}\right) - \log\left(R_{1,t}\right) = \lim_{S_{0,0}\to\infty} \log\left(R_{0,t+1}\right) - \log\left(R_{0,t}\right) \iff$$

$$\lim_{S_{1,0}\to\infty} \log\left(\frac{R_{1,t_2}}{R_{1,t_1}}\right) = \lim_{S_{0,0}\to\infty} \log\left(\frac{R_{0,t_2}}{R_{0,t_1}}\right) \iff$$

Substituting from Proposition 9 $\longrightarrow$

$$\lim_{S_{1,0}\to\infty} \log\left(\frac{\sum_{j=t_2}^{\infty} (1-\gamma)^{j-t_2}\, \beta_{1,j} S_{1,j}}{\sum_{j=t_1}^{\infty} (1-\gamma)^{j-t_1}\, \beta_{1,j} S_{1,j}}\right) = \lim_{S_{0,0}\to\infty} \log\left(\frac{\sum_{j=t_2}^{\infty} (1-\gamma)^{j-t_2}\, \beta_{0,j} S_{0,j}}{\sum_{j=t_1}^{\infty} (1-\gamma)^{j-t_1}\, \beta_{0,j} S_{0,j}}\right) \iff$$

Taking limits $\longrightarrow$

$$\log\left(\frac{\sum_{j=t_2}^{\infty} (1-\gamma)^{j-t_2}\, \beta_{1,j}}{\sum_{j=t_1}^{\infty} (1-\gamma)^{j-t_1}\, \beta_{1,j}}\right) = \log\left(\frac{\sum_{j=t_2}^{\infty} (1-\gamma)^{j-t_2}\, \beta_{0,j}}{\sum_{j=t_1}^{\infty} (1-\gamma)^{j-t_1}\, \beta_{0,j}}\right)$$

□

Parallel trends ($\log(\beta_t)$)

Proposition (Parallel trends: Log β_t)

Assuming an SIR data-generating process and a log-transformed effective reproduction number specification $(Y_{d,t} = \log(\beta_{d,t}), g(\cdot) = \log(\cdot))$*, the parallel trends assumption holds for all* $t_1, t_2 > t-1$ *if and only if*

$$\log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$
$$\log\left(\beta_{1,t_2}\right) - \log\left(\beta_{1,t_1}\right) = \left(\log(\beta_{0,t_2}\right) - \log\left(\beta_{0,t_1}\right)$$

Return

ATT Interpretations

Specification	**Outcome**	**Link** $g(.)$	**Interpretation of ATT**
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	Difference
Log incidence	$Y_{d,t} = I^*_{d,t}$	log	Percentage difference
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left(I^*_{d,t}\right)}{\mathbb{E}\left(I^*_{d,t-1}\right)}$	log	Percentage change
Log R_t	$Y_{d,t} = R_{d,t}$	log	Difference in avg. transmissions per infection
Log β_t	$Y_{d,t} = \beta_{d,t}$	log	Difference in effective contact rate

Return

Average marginal effects

Algorithm 1 (Estimation of average marginal effects for log incidence and log growth specifications) Given the observed outcomes in the treated units, $Y_1, \ldots, Y_{N_1}$, and the estimated ATT, $\hat{\delta}_t$, we impute the AME:

1. Calculate the fitted untreated potential outcome for the treated group in the scale of the model specification for each treated unit i and post-intervention period $t > T_0$ using the observed empirical outcome trajectory and the estimated ATT: $\widehat{Y_d}(0) = Y_{d,t} - \hat{\delta}_t$
2. Recover the fitted untreated potential outcome for the treated unit i in the case scale, $\widehat{I^*_{d,t}}(0)$, from $\hat{Y}_{d,t}(0)$ according to the definition of model specifications per Table **??**. For log growth, we take the last period prior to intervention as baseline, and construct the untreated potential outcomes by dividing the baseline outcome by the fitted treatment effect coefficient. We repeat division for each post-intervention period to recover the untreated trajectories for the treated units.
3. Calculate the difference between the observed treated outcome and the fitted control potential outcome trajectories to obtain the marginal effect (ME) for each unit i over the entire post-intervention time periods: $ME_i = \sum_{t=T_0+1}^{T} \left(I^*_{d,t} - \widehat{I^*_{d,t}}(0) \right)$
4. The AME is given by the average of the calculated differences over all treated units: $AME = \frac{1}{N_1} \sum_{i=1}^{N} (ME_i)$

Algorithm 2 (Estimation of average marginal effects for log R_t or log β_t models) For COVID-19, we assume on average 5 days of infectiousness and 3 days of mean exposure period. We use input data on the initial susceptible fraction and infections, as well as empirically estimated effective contact rates over the period of interest $\beta_t, t \in [t_1, t_2]$. We then use estimated time-varying ATTs for the effective contact rate, $\hat{\delta}_t$ to impute the AME:

1. Calculate the fitted treated potential outcomes as an average from 1000 infection trajectories simulated from an SEIR model with effective contact rates set to $\left(\beta_t + \hat{\delta}_t\right)$, corresponding to an effective reproduction number $R_t = 5\left(\beta_t + \hat{\delta}_t\right)$.
2. Calculate the fitted untreated potential outcomes for the treated group as an average from 1000 infection trajectories simulated from an SEIR model with effective contact rate set to β_0, corresponding to an effective reproduction number $R_t = 5\beta_0$.
3. The AME for a log R_t or a log β_t model is then given by the difference in projected infections over the post-intervention period between the two fitted trajectories.

Return

Inference

We conduct inference using the wild score bootstrap, which allows for valid inference with heteroskedastic data and a small number of clusters when a generalized linear model is used for estimation. This is a generalization of the wild cluster bootstrap. Given any maximum likelihood estimation process, in each bootstrap replicate,

1. Estimate the score contribution for cluster c as the sum of score vectors in all observations from cluster c, where a score vector is the first derivative of the log-likelihood function.
2. Re-weight the score distribution based on an auxiliary cluster-level random variable with mean 0 and variance 1.
3. Calculate a Wald statistic is calculated using the weighted scores.

The p-value is the proportion of bootstrap replicates for which the bootstrapped Wald statistics exceed the observed Wald statistic under the null.

Return

R_t and β_t estimation (SIR)

Proposition (Estimation of $\beta_{d,t}$)

Assuming an SIR data-generating process, with $I_{d,t+1} \sim Pois\left(\beta_{d,t} S_{d,t} \frac{S_{d,t}}{N}\right)$, *the maximum likelihood estimator of* $\beta_{d,t}$ *is:*

$$\hat{\beta}_{d,t} = \frac{I^*_{d,t+1}}{I_{d,t} \frac{S_{d,t}}{N}}$$

Proof.

Because we assume:

$$I^*_{t+1} | I_t \sim Pois\left(\beta_t I_t \frac{S_t}{N}\right),$$

the log-likelihood (ℓ) function can be defined:

$$\ell\left(\beta_t | I^*_{t+1}, I_t, S_t, N\right) \propto I^*_{t+1} \log\left(\beta_t I_t \frac{S_t}{N}\right) - \beta_t I_t \frac{S_t}{N}$$

Setting $\frac{\partial \ell\left(\beta_t | I^*_{t+1}, I_t, S_t, N\right)}{\partial \beta_t} = 0$ to obtain the maximum likelihood estimator:

$$0 = \frac{\ell\left(\beta_t | I^*_{t+1}, I_t, S_t, N\right)}{\partial \beta_t}$$

$$\implies \hat{\beta}_t = \frac{I^*_{t+1}}{I_t \frac{S_t}{N}}$$

□

Timestep aggregation

Proposition (DiD with time-step aggregation)

Suppose that the parallel trends assumption holds with a log link for every pair of individual pre- and post-intervention time steps between the average outcome in the treated and comparison groups. That is, for any individual time steps $t_1 \leq T_0$ and $T_0 < t_2 \leq T$, we assume

$$\log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right) = \log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right)$$

Then, the following parallel trends assumption holds over aggregated time intervals:

$$\log\left(\mathbb{E}\left[Y_{1,t}(0)\middle| t \in \tau_2\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t}(0)\middle| t \in \tau_1\right]\right) = \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle| t \in \tau_2\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle| t \in \tau_1\right]\right),$$

where τ_1 and τ_2 denote arbitrary combination of pre- and post-intervention time periods, respectively.

Return

Multiple units and time periods

We can extend the parallel trends assumption to both multiple units and multiple time periods:

$$g\Big(\mathbb{E}\left[Y_{i,t}(0)\middle|i \in \mathcal{N}_1, t \in \mathcal{T}_1\right]\Big) - g\Big(\mathbb{E}\left[Y_{i,t}(0)\middle|i \in \mathcal{N}_1, t \in \mathcal{T}_0\right]\Big) =$$
$$g\Big(\mathbb{E}\left[Y_{i,t}(0)\middle|i \in \mathcal{N}_0, t \in \mathcal{T}_1\right]\Big) - g\Big(\mathbb{E}\left[Y_{i,t}(0)\middle|i \in \mathcal{N}_0, t \in \mathcal{T}_0\right]\Big), \tag{1}$$

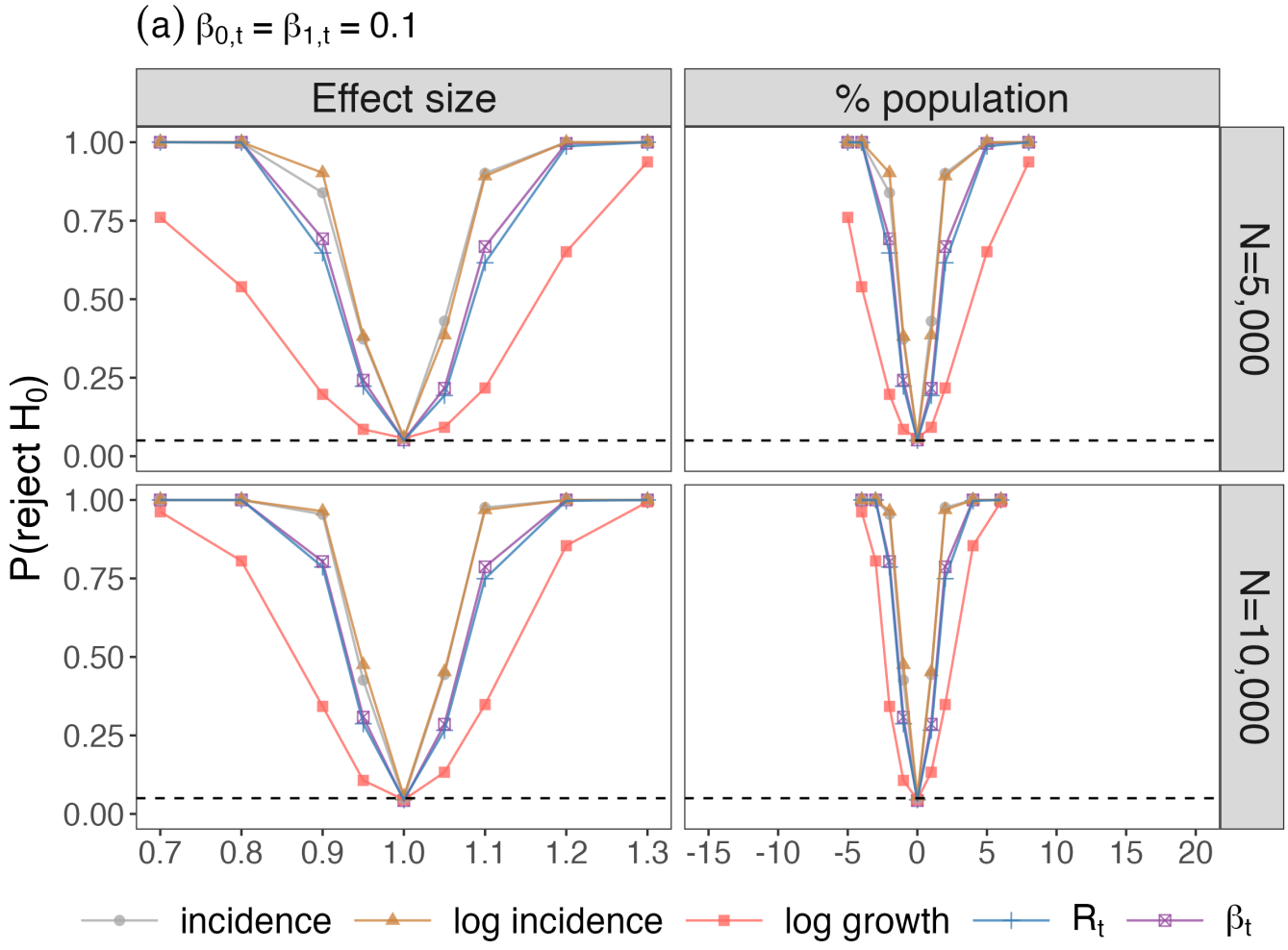
- Multiple periods: sufficient to assume parallel trajectories
- Multiple units: more complex for incidence

Return

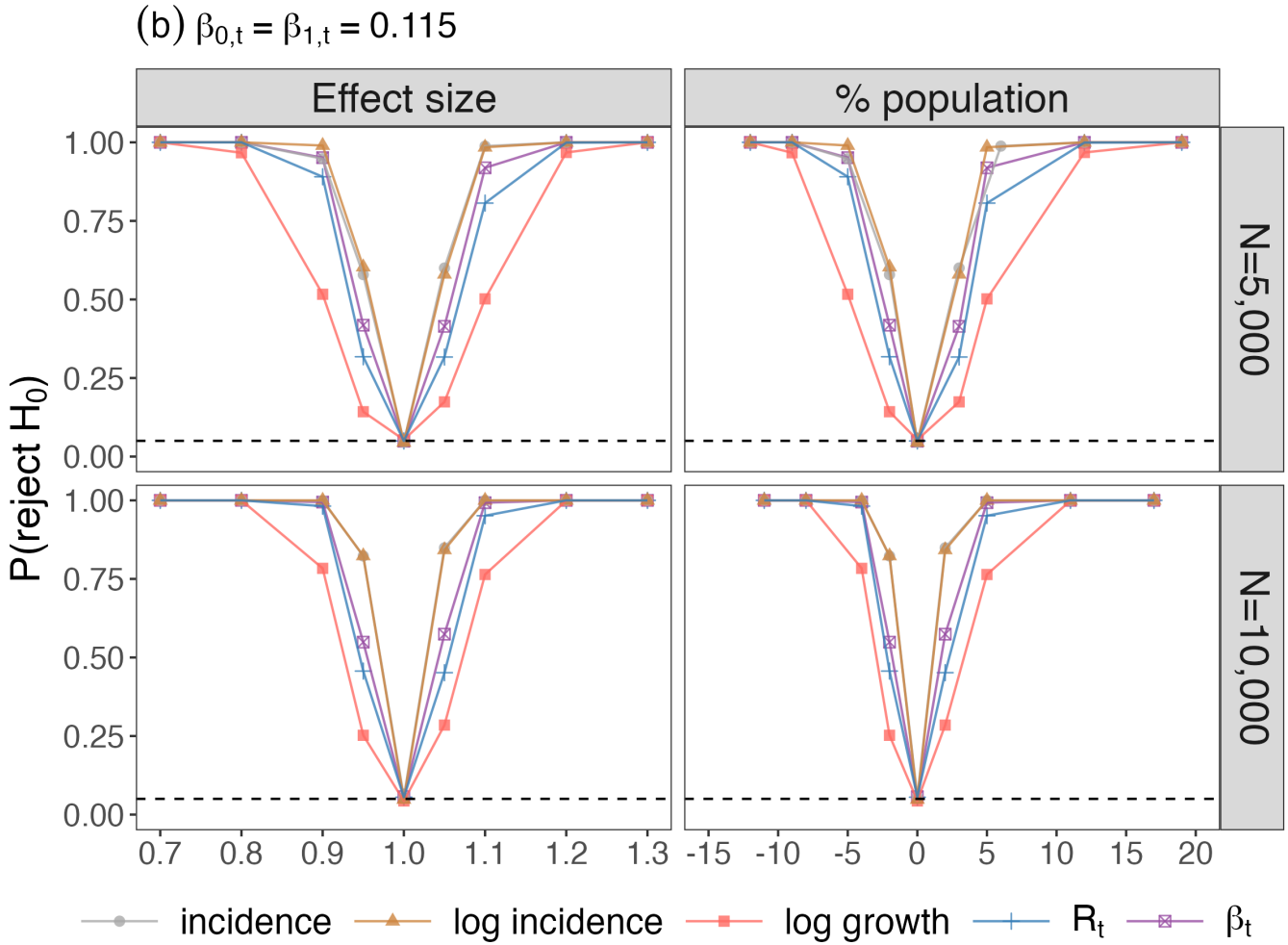
Power simulation parameters

Parameter	Description	Values
N	Total number of units	50
N_1	Number of treated units	25
pop	Population size for each unit	$\{5,000,\ 10,000\}$
T	Total number of weeks	17
T_{burnin}	Weeks in the burn-in period	5
T_0	Weeks in the pre-intervention period	4
I_0	Infections at time 0	100
β	Effective contact rate	$\{0.100,\ 0.115\}$
γ	Generation interval	10 days
ϕ	Ratio in the effective contact rates between the treated and control groups	$\{1.0,\ 1.1\}$
δ	Effect size	$\{0.70, 0.80, 0.90, 0.95, 1.00, 1.05, 1$
α	Statistical significance level	0.05

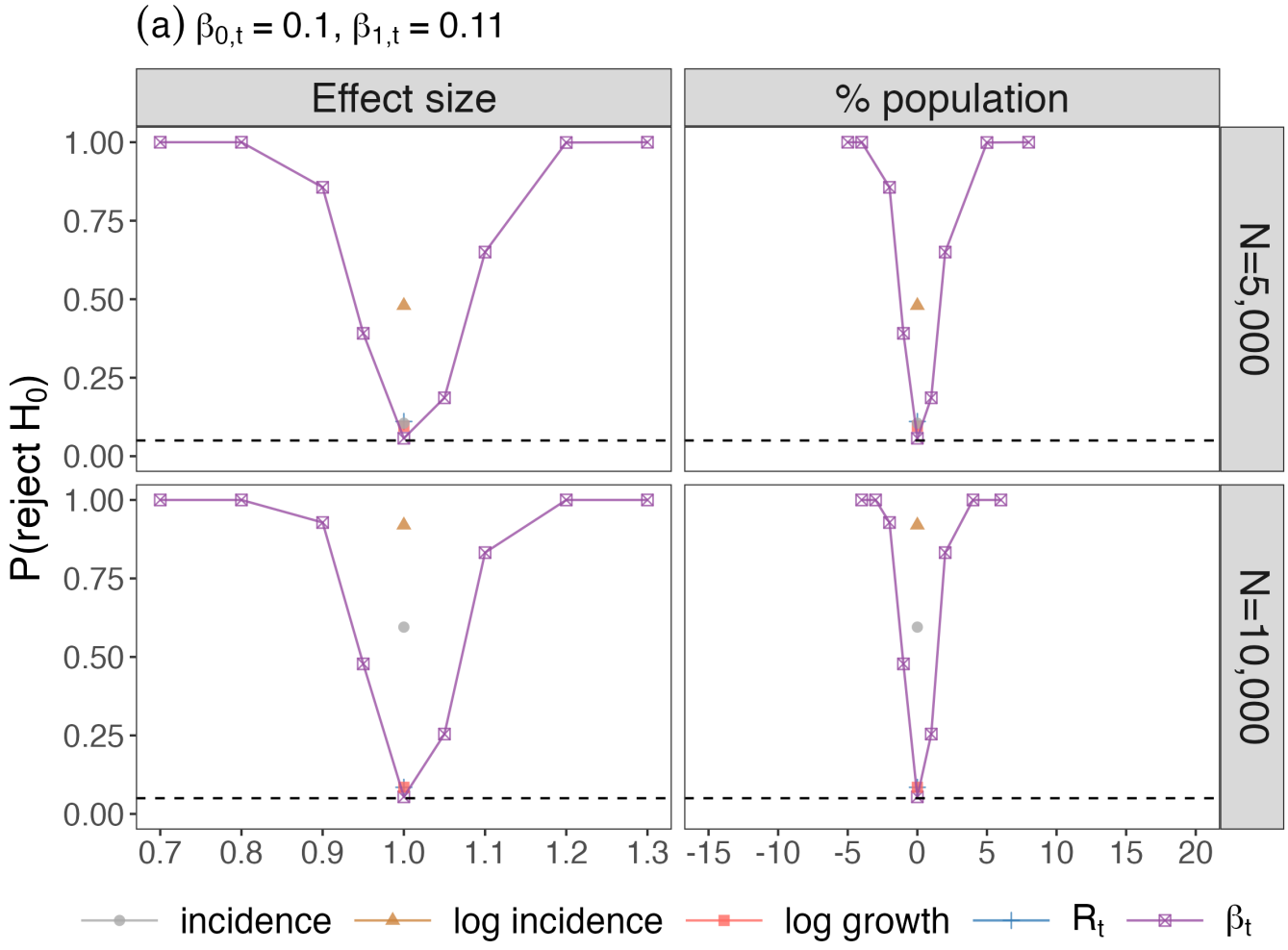
Return


(a) $\beta_{0,t} = \beta_{1,t} = 0.1$
Effect size
% population
N=5,000
N=10,000
P(reject H_0)
1.00
0.75
0.50
0.25
0.00
0.7
0.8
0.9
1.0
1.1
1.2
1.3
-15
-10
-5
0
5
10
15
20
incidence
log incidence
log growth
R_t
β_t

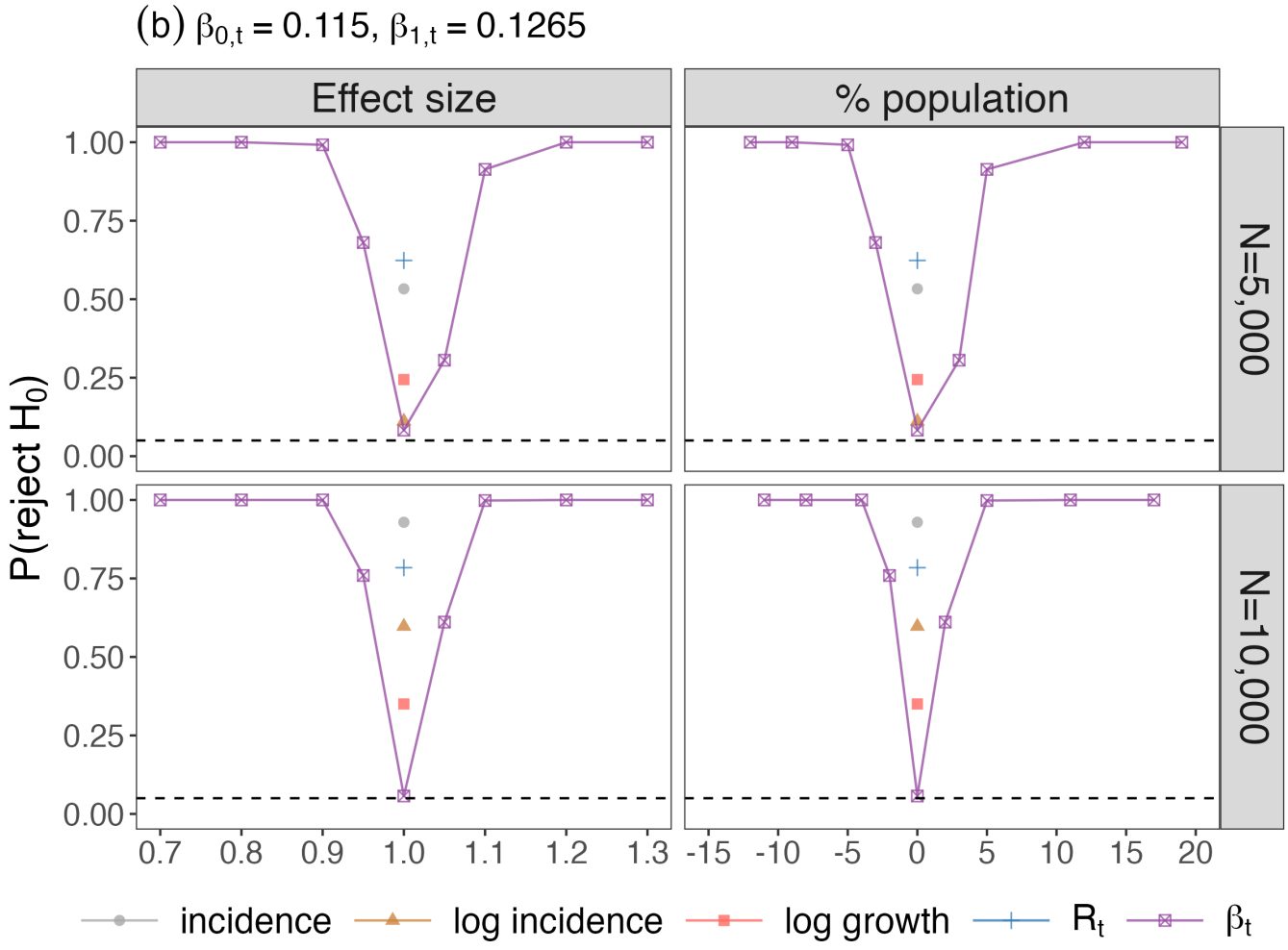
Return


(b) $\beta_{0,t} = \beta_{1,t} = 0.115$
Effect size
% population
N=5,000
N=10,000
P(reject H_0)
1.00
0.75
0.50
0.25
0.00
0.7
0.8
0.9
1.0
1.1
1.2
1.3
-15
-10
-5
0
5
10
15
20
incidence
log incidence
log growth
R_t
β_t

Return


(a) $\beta_{0,t} = 0.1$, $\beta_{1,t} = 0.11$
Effect size
% population
N=5,000
N=10,000
P(reject H_0)
1.00
0.75
0.50
0.25
0.00
0.7
0.8
0.9
1.0
1.1
1.2
1.3
-15
-10
-5
0
5
10
15
20
incidence
log incidence
log growth
R_t
β_t

Return


(b) $\beta_{0,t} = 0.115$, $\beta_{1,t} = 0.1265$
Effect size
% population
N=5,000
N=10,000
P(reject H_0)
1.00
0.75
0.50
0.25
0.00
0.7
0.8
0.9
1.0
1.1
1.2
1.3
-15
-10
-5
0
5
10
15
20
incidence
log incidence
log growth
R_t
β_t

Return