Parallel Trends in an Unparalleled Pandemic Difference-in-differences for infectious disease policy evaluation

Shuo Feng[a] **and Alyssa Bilinski, PhD**[b,1]

April 1, 2024

Researchers frequently employ difference-in-differences (DiD) to study the impact of public health interventions on infectious disease outcomes. DiD assumes that treatment and non-experimental comparison groups would have moved in parallel in expectation, absent the intervention ("parallel trends assumption"). However, the plausibility of parallel trends assumption in the context of infectious disease transmission is not well-understood. Our work bridges this gap by formalizing epidemiological assumptions required for common DiD specifications, positing an underlying Susceptible-Infectious-Recovered (SIR) data-generating process. We demonstrate that popular specifications can encode strict epidemiological assumptions. For example, DiD modeling incident case numbers or rates as outcomes will produce biased treatment effect estimates unless untreated potential outcomes for treatment and comparison groups come from a data-generating process with the same initial infection and equal transmission rates at each time step. Applying a log transformation or modeling log growth allows for different initial infection rates under an "infinite susceptible population" assumption, but invokes conditions on transmission parameters. We then propose alternative DiD specifications based on epidemiological parameters – the effective reproduction number and the effective contact rate – that are both more robust to differences between treatment and comparison groups and can be extended to complex transmission dynamics. With minimal power difference incidence and log incidence models, we recommend a default of the more robust log specification. Our alternative specifications have lower power than incidence or log incidence models, but have higher power than log growth models. We illustrate implications of our work by re-analyzing published studies of COVID-19 mask policies.

Difference-in-differences | Transmission dynamics | Infectious disease models | Observational causal inference

Significance Statement

Difference-in-differences is a popular observational study design for policy evaluation. However, it may not perform well when modeling infectious disease outcomes. Although many COVID-19 DiD studies in the medical literature have used incident case numbers or rates as the outcome variable, we demonstrate that this and other common model specifications may encode strict epidemiological assumptions as a result of non-linear infectious disease transmission. We unpack the assumptions embedded in popular DiD specifications assuming a Susceptible-Infected-Recovered data-generating process and propose more robust alternatives, modeling the effective reproduction number and effective contact rate.

Throughout the COVID-19 pandemic, researchers extensively studied the impact of public health interventions on disease incidence and mortality, most often with observational study designs (1). Difference-in-differences (DiD), already widely-used in health and social sciences, was one popular approach for this work. DiD assumes that treatment and non-experimental comparison groups would have moved in parallel in expectation absent the intervention (the "parallel trends assumption"). It uses this functional form assumption to impute counterfactual potential outcomes and estimate treatment effects. In the context of COVID-19, DiD was employed to evaluate policies including social distancing (2), school reopening (3), stay-at-home orders (4), and school mask mandates (5) in the United States, as well mask mandates in Germany (6) and contact tracing in England (7).

Outside of infectious disease, DiD has traditionally been used with outcomes that are expected to evolve linearly over time. By contrast, epidemiological theory predicts that infectious pathogens will spread non-linearly as a function of interactions between susceptible and infectious individuals. Mechanistic transmission dynamic models that capture these interactions are often used in epidemiology to prospectively project the impact of potential disease mitigation measures. However, these models typically rely on transporting effects from mechanistic or small-scale studies, and there are growing concerns that this may produce overly optimistic estimates of the impact of policies and programs (8). For example, randomized controlled trial estimates of the population-level reduction in transmission from mask mandates were smaller than those used to inform some transmission dynamic estimates (9). These limitations

Author affiliations: [a]Department of Biostatistics, Brown University School of Public Health; [b]Departments of Health Services, Policy & Practice and Biostatistics, Brown University School of Public Health

S.F. and A.M.B. designed and performed research; S.F. analyzed data; and S.F. and A.M.B. wrote the manuscript.

The authors declare no competing interests.

[1] Corresponding author: alyssa_bilinski@brown.edu

have increased interest in post-hoc policy evaluation using methods like DiD to compare outcomes in treated areas with those from similar untreated units.

However, there remains insufficient guidance on how to best account for the quasi-exponential infectious disease dynamics in DiD. Prior DiD work has noted that the parallel trends assumption may hold for some model specifications but not others (10, 11). In particular, Roth and Sant'Anna (10) demonstrated that the parallel trends assumption is sensitive to functional form unless strict conditions are met: if the population can be divided into two groups, one with random treatment assignment and another with stable distribution of the untreated potential outcomes over both pre- and post-intervention periods. Others have noted trade-offs between robustness and power and suggested applying domain-specific theory to select functional form (11).

Infectious disease transmission dynamic theory is well-developed, offering an opportunity to guide DiD specification. Nevertheless, one systematic review found that less than one-fifth of COVID-19 health policy evaluations, including many DiD applications, justified functional form or their choice of model specifications (12). To better understand the use of DiD in recent COVID-19 literature, we conducted a comprehensive review of all COVID-19 DiD analyses published in *Journal of the American Medical Association (JAMA)* network journals, *New England Journal of Medicine (NEJM)*, *Proceedings of the National Academy of Sciences (PNAS)*, *Nature Research* journals, *Lancet* journals, *Health Affairs*, and *Health Economics* from 2020-2022 (Appendix A). We observed considerable variation in their model specifications. Most publications (17 of the 29 papers reviewed, 59%) used incident counts or rates of cases or deaths as outcomes. Another 10 (34%) considered log-transformed incidence as the outcome measure. The remaining 2 studies (7%) specified the growth rate in log incidence as an outcome measure.

Most papers we analyzed (over 85%) did not offer specific epidemiological justification for their outcome choice. However, some researchers, particularly in economics, have cautioned against modeling incidence or mortality directly in DiD with infectious disease outcomes (3, 4). For example, Callaway and Li (4) noted that with an SIR data-generating process, the parallel trends assumption would not hold if untreated treatment and comparison trajectories differed and proposed an unconfoundedness approach to condition on pre-treatment state, in the spirit of matching on the full pre-treatment pandemic trajectory. Two papers using log growth rate as their outcome measures also motivated their specification with an SIR model (3, 13).

We extend this prior work along three key dimensions. First, we comprehensively catalog epidemiological assumptions required for DiD to recover unbiased treatment effects with different model specifications, assuming an underlying SIR process. This both summarizes the literature from an epidemiological perspective and fills in gaps by, for example, linking the log model specification to specific epidemiological parameters and expanding prior discussion of log growth models to address susceptible depletion. Second, we propose new model specifications based on epidemiological parameters. We show that these can both recover unbiased treatment effects under less strict assumptions and be applied to more complex transmission dynamic processes. Finally, we explore the power of different model specifications and highlight trade-offs involved in employing more robust specifications.

The rest of the paper proceeds as follows. Section 2 provides an overview of SIR and DiD models. In Section 3, we characterize epidemiological assumptions required for DiD to produce unbiased treatment effects assuming an underlying SIR data-generating process. We then propose alternative DiD model specifications, based on parallel trends in the effective reproduction number or effective contact rate. Section 4 explores the statistical power of different model specifications. Last, Section 5 demonstrates the implications of our findings by re-analyzing previously published work on evaluating the effect of mask policies on COVID-19 cases.

Models

Susceptible-Infectious-Recovered (SIR). Transmission dynamic models assume that infectious diseases spread based on the frequency and intensity of interactions between susceptible and infectious individuals. Susceptible-Infectious-Recovered (SIR) models are a popular class of models that assume exponential growth at the start of an outbreak, with incidence declining as the susceptible population is depleted.

For this work, we posit a stochastic SIR data-generating process that incorporates randomness in transmission and, extending some prior work (e.g., Callaway and Li (4)), allows transmission intensity to vary over time, reflecting shifts in precautionary behaviors, vaccination, and variants. Given initial populations of susceptible, infectious, and recovered individuals in group d, $\{S_{d,0},\ I_{d,0},\ R_{d,0}\}$, we assume the number of individuals in each state evolves according to the following set of equations:

$$\begin{aligned}
&\textbf{Incidence: } I^*_{d,t+1} \sim Pois\left(\mu_t = \beta_{d,t} I_{d,t} \frac{S_{d,t}}{N}\right) \qquad [1]\\
&\textbf{Susceptible: } S_{d,t+1} = S_{d,t} - I^*_{d,t}\\
&\textbf{Infectious: } I_{d,t+1} = (1-\gamma) I_{d,t} + I^*_{d,t}\\
&\textbf{Recovered: } R_{d,t+1} = R_{d,t} + \gamma I_{d,t}
\end{aligned}$$

In this setup, we denote the number of incident infections for unit d at time t as $I^*_{d,t+1}$ (in contrast to prevalent infections, $I_{d,t+1}$). Incident infections at time $t+1$ depend on $\beta_{d,t}$, which we denote the effective contact rate; the size of the infected population, $I_{d,t}$; and the fraction of susceptible individuals, $S_{d,t}/N$, all in the previous time step, t. Per standard practice, we assume incidence follows a Poisson distribution, although most results presented can accommodate other distributions with a mean of μ_t (4)*. The number of removed (recovered or dead) individuals at time t is $\gamma I_{d,t}$, where $\frac{1}{\gamma}$ is the average length of the infectious period, equal to the generation interval in an SIR framework, i.e., the time from the infection of a primary case to a secondary infection generated.

In practice, researchers are often interested in contexts in which the susceptible population remains large relative to active and recovered infections over the period of interest, meaning that $\mathbb{E}\left[I^*_{d,t}\right] \approx \beta_{d,t} I_{d,t}$. We formalize this mathematically as an "infinite susceptible population" scenario, the limit of the data-generating process as $S_{d,0} \to \infty$.

SIR models assume a closed and stable population of N individuals for every unit d with homogeneous mixing

*Astute readers may note that it is possible for $I^*_{d,t}$ to exceed $S_{d,t}$. This rarely occurs as typically $\mu_t << S_{d,t}$, but if a concern, the Poisson distribution can be substituted with a binomial.

(i.e., $\beta_{d,t} I_{d,t}/N$ to be constant across individuals). Although these assumptions are unlikely to hold exactly, they may be reasonable approximations in many applications, when researchers model outcomes at an aggregated level (e.g., counties or states). However, they may be less plausible with more granular units of analysis or when the population size is changing rapidly. We analyze this simple SIR process in closed-form; in **Extensions**, we discuss generalizations to more complex transmission dynamics.

Difference-in-Differences (DiD). We begin with a canonical DiD setup with two units $d \in \{0,1\}$ and two time periods $t \in \{t_1, t_2\}$. Assume unit 1 is treated at time t_2 and unit 0 remains untreated in both time periods (14). We denote the observed outcome of interest for unit d at time t as $Y_{d,t}$ and the potential outcome for unit d at time t given treatment $k \in \{0,1\}$ as $Y_{d,t}(k)$. The average treatment effect on the treated (ATT) at time t_2 is the expected difference in potential outcomes for unit 1 at time t_2 under intervention and no-intervention scenarios:

$$ATT = \mathbb{E}\Big(Y_{1,t_2}(1) - Y_{1,t_2}(0)\Big)$$

Because $Y_{1,t_2}(0)$ is unobservable, DiD relies on the parallel trends assumption, that the untreated potential outcomes in both units evolve in parallel from t_1 to t_2, i.e.,

$$\mathbb{E}\Big(Y_{1,t_2}(0) - Y_{1,t_1}(0)\Big) = \mathbb{E}\Big(Y_{0,t_2}(0) - Y_{0,t_1}(0)\Big) \quad [2]$$

Under this assumption, we can estimate an unbiased treatment effect by solving for $Y_{1,t_2}(0)$ in Eq. 2 and applying sample analogs:

$$\widehat{ATT} = \big(Y_{1,t_2} - Y_{1,t_1}\big) - \big(Y_{0,t_2} - Y_{0,t_1}\big)$$

We can extend the parallel trends assumption to monotonic, continuously differentiable transformations ($g(\cdot)$) (15):

$$g\Big(\mathbb{E}\left[Y_{1,t_2}(0)\right]\Big) - g\Big(\mathbb{E}\left(Y_{1,t_1}(0)\right)\Big) = g\Big(\mathbb{E}\left[Y_{0,t_2}(0)\right]\Big) - g\Big(\mathbb{E}\left[Y_{0,t_1}(0)\right]\Big) \quad [3]$$

In the following sections, we establish conditions under which the transformed non-linear version of parallel trends assumption (Eq. 3) holds under different specifications of the DiD model, i.e., different $Y_{d,t}$ and $g(\cdot)$. In **Extensions**, we generalize findings to DiD applications with more than two units or time periods, noting that the conditions remain largely unchanged.

Model specifications

Assuming an underlying SIR data-generating process, different model specifications encode different epidemiological conditions required for the parallel trends assumption to hold, allowing unbiased estimation of the ATT with DiD. In this section, we derive and provide intuition for assumptions invoked across various specifications.

We summarize model specifications considered in Table 1, increasing in robustness. These include the three specifications identified in our literature review, as well as two new proposed specifications modeling the effective reproduction number and effective contact rate. Each specification implies a different interpretation of the ATT as noted in the fourth column of Table 1. As a result, we detail methods for imputing the average marginal effect on the incidence scale in **Average marginal effects**.

For log specifications in Table 1, we assume that the parallel trends assumption is defined in terms of $\log\left(\mathbb{E}\left[Y_{d,t}\right]\right)$. However, 8 (or 28%) papers modeled log specifications by transforming the outcome variable, assuming parallel trends in $\mathbb{E}\left[\log\left(Y_{d,t}\right)\right]$, rather than $\log\left(\mathbb{E}\left[Y_{d,t}\right]\right)$. Although $\mathbb{E}\left[\log\left(Y_{d,t}\right)\right] \neq \log\left(\mathbb{E}\left[Y_{d,t}\right]\right)$ (Jensen's inequality), the two are nearly equivalent when $Y_{d,t}$ is sufficiently large (Supplement Proposition 1 and Corollary 1), and in practice, results are unlikely to be affected by this approximation.

Last, note that each model specification in Table 1 can be formulated based either on the count or rate per population of infections. In the literature reviewed (Appendix A), about half (48%) of researchers modeled rates (i.e., infections or cases per unit population) rather than numbers of infections or cases. This assumes frequency-dependent transmission, i.e., the average number of secondary infections per infectious individual remains relatively constant across different population sizes (16). For parsimony in this section, we assume that units have equal population sizes, obviating the need to scale in derivations, and results that follow could apply either to frequency-dependent or density-dependent transmission. However, if population sizes differ and frequency-dependent transmission is assumed, derivations can be adapted accordingly by scaling model specifications by population size (e.g., incidence per 100,000 population rather than incident infection numbers).

Established model specifications.

Incidence. As previously noted, incidence was the most popular model specification: $Y_{d,t} = I^*_{d,t}$ (with the identity link function). To understand epidemiological assumptions embedded in DiD with incidence outcomes, we first obtain an expression for expected incidence.

Proposition 1 (Expected incidence)**.** *Assuming an SIR data-generating process (Eq. 1) with initial conditions* $\{S_{d,0}, I_{d,0}, R_{d,0}\}$*, expected incidence at time* $t+1$ *can be written as:*

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

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

We derive Proposition 1 in Appendix B. Even with t_1 and t_2 as adjacent time-steps, we cannot write $\mathbb{E}\left[I^*_{d,t+1}\right]$ as a linear function of $\mathbb{E}\left[I^*_{d,t}\right]$. Indeed, with this data-generating process, there are not straightforward conditions under which the parallel trends assumption holds for incidence, log incidence, or log growth models other than equality in all data-generating parameters.

We therefore explore assumptions required under "infinite susceptible population" conditions (i.e., $S_{d,0} \to \infty$), implying that the susceptible population is very large relative to active and recovered infections. In this case, expected incidence can be written as an iterative product of prior effective contact rates and initial infection.

Proposition 2 (Expected incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1), with initial conditions* $\{S_0, I_0, R_0\}$*, for* $t \geq 1$*,*

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

We then solve for conditions under which the "infinite susceptible population" parallel trends assumption holds as $S_{1,0}, S_{0,0} \to \infty$. Because $E\left[\lim_{S_{d,0}\to\infty} I^*_{d,t}\right] = \lim_{S_{d,0}\to\infty} E\left[I^*_{d,t+1}\right]$, and $g(.)$ is continuous, we write $\lim_{S_{d,0}\to\infty} g\left(\mathbb{E}\left[Y_{d,t}\right]\right)$ to compress notation; because units are assumed independent, taking the limit as $S_{1,0}, S_{0,0} \to \infty$ is equivalent to taking the corresponding limit on each unit.

Proposition 3 (Parallel trends: Incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) and an incidence model specification* $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\mathbb{E}\left[Y_{1,t_2}(0)\right] - \mathbb{E}\left[Y_{1,t_1}(0)\right]\right) =$$
$$\lim_{S_{0,0}\to\infty}\left(\mathbb{E}\left[Y_{0,t_2}(0)\right] - \mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff$$
$$\mathbb{E}\left[I^*_{1,t_1}\right]\left(\beta^*_{1,t_1,t_2} - 1\right) = \mathbb{E}\left[I^*_{0,t_1}\right]\left(\beta^*_{0,t_1,t_2} - 1\right),$$
$$\text{where } \mathbb{E}\left[I^*_{d,t_1}\right] = I_{d,0}\beta_{d,t_1-1}\prod_{k=0}^{t_1-2}(1-\gamma+\beta_{d,k}),$$
$$\beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}}\prod_{k=t_1-1}^{t_2-2}(1-\gamma+\beta_{d,k})$$

Proposition 3 follows from substituting results of Proposition 2 into Eq. 3 (proof in Appendix B). This result highlights that the parallel trends assumption imposes strong conditions on underlying transmission dynamics, even assuming an "infinite susceptible population. Proposition 3 holds when groups start with the same expected incidence $\left(\mathbb{E}\left[I^*_{d,t_1}\right]\right)$ at time t_1 and continue to grow with a set of time-varying effective contact rates (β_t) and generation interval (γ) that produce equal β^*_{d,t_1,t_2}. (Parallel trends over time would require equality across groups in this set of effective contact rates). We illustrate these assumptions the first row of Figure 1, where trends are only parallel (noted by a checkmark) when all input parameters are equal, and as a result expected incidence trajectories match, between treatment and comparison groups.

Proposition 3 also implies that traditional DiD diagnostics like event study plots may be misleading for assessing the plausibility of parallel trends assumption. In contrast to traditional DiD which allows for level differences, this specification requires evidence of equal epidemic trajectories. Furthermore, visual inspections may be misleading at the start of a new wave of disease, as groups may initially appear similar, but increasingly diverge as a wave accelerates even absent a treatment effect (Figure 1, panel 1(b)).

Log incidence. The second most common model specification was log incidence, $Y_{d,t} = I^*_{d,t}$ with $g(\cdot) = \log(\cdot)$, often motivated by the idea that disease transmission can be exponential or nearly exponential (17, 18).

Proposition 4 (Parallel trends: Log incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) 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 (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

$$\lim_{S_{1,0}\to\infty}\left(\log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right)\right) =$$
$$\lim_{S_{0,0}\to\infty}\left(\log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right)\right) \iff$$
$$\beta^*_{1,t_1,t_2} = \beta^*_{0,t_1,t_2},$$
$$\text{where } \beta^*_{d,t_1,t_2} = \frac{\beta_{d,t_2-1}}{\beta_{d,t_1-1}}\prod_{k=t_1-1}^{t_2-2}(1-\gamma+\beta_{d,k})$$

The derivation for Proposition 4 follows similar logic as that of Proposition 3 (Appendix B). Per Proposition 4, with an "infinite susceptible population," when log incidence is used as outcome, the parallel trends assumption no longer imposes restrictions on expected incidence at time t_1. Nevertheless, it still requires an equal product of effective contact rates in both units between t_1 and t_2. We illustrate these conditions in the second row of Figure 1.

Log growth. The third approach identified in our literature review ("log growth") modeled the change in log incidence over adjacent time steps. Because this ratio is undefined in the context where 0 is in the support of $I^*_{d,t}$, we model this as $Y_{d,t} = \frac{\mathbb{E}\left[I^*_{d,t}\right]}{\mathbb{E}\left[I^*_{d,t-1}\right]}$, with $g(\cdot) = \log(\cdot)$.[†]

Proposition 5 (Parallel trends: Log growth (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) 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 (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

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

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

[†] This framing is consistent with the practice of collapsing multiple time periods into a single step to address zero-valued outcomes when using log growth models.

Table 1. Transmission dynamic assumptions required for the parallel trends assumption, assuming an SIR data-generating process.

Specification	Outcome	Link[1] $g(.)$	Interpretation of ATT	Assumptions: Treatment vs. comparison parameters		
				Susceptible population[2] $(S_{d,0})$	Initial incidence (Y_{d,t_1})	Effective contact rates $(\beta_{d,t_1-1},...,\beta_{d,t_2-1})$
Incidence	$Y_{d,t} = I^*_{d,t}$	identity	Difference	$=$ or $S_{d,0} \to \infty$	$=$	$=$
Log incidence	$Y_{d,t} = I^*_{d,t}$	log	Percentage difference	$S_{d,0} \to \infty$		$=$
Log growth	$Y_{d,t} = \frac{\mathbb{E}\left(I^*_{d,t}\right)}{\mathbb{E}\left(I^*_{d,t-1}\right)}$	log	Percentage change	$S_{d,0} \to \infty$		constant ratio + $\beta_{d,t} = \beta_d \ \forall t$ or $\gamma = 1$
Log R_t[3]	$Y_{d,t} = R_{d,t}$	log	Difference in avg. transmissions per infection	$S_{d,0} \to \infty$	known	constant ratio
Log β_t[3]	$Y_{d,t} = \beta_{d,t}$	log	Difference in effective contact rate		known	constant ratio

[1] Log links can also be implemented as a transformation of $Y_{d,t}$ (e.g. $Y_{d,t} = \log(I^*_{d,t})$, with minimal bias provided the outcome value is sufficiently large (Supplement Proposition 1). [2] For the first four specifications, we list assumptions provided that $S_{d,0} \to$, $\forall d$. Without this assumption, we require stronger conditions on the full set of parameters. [3] Requires known generation interval

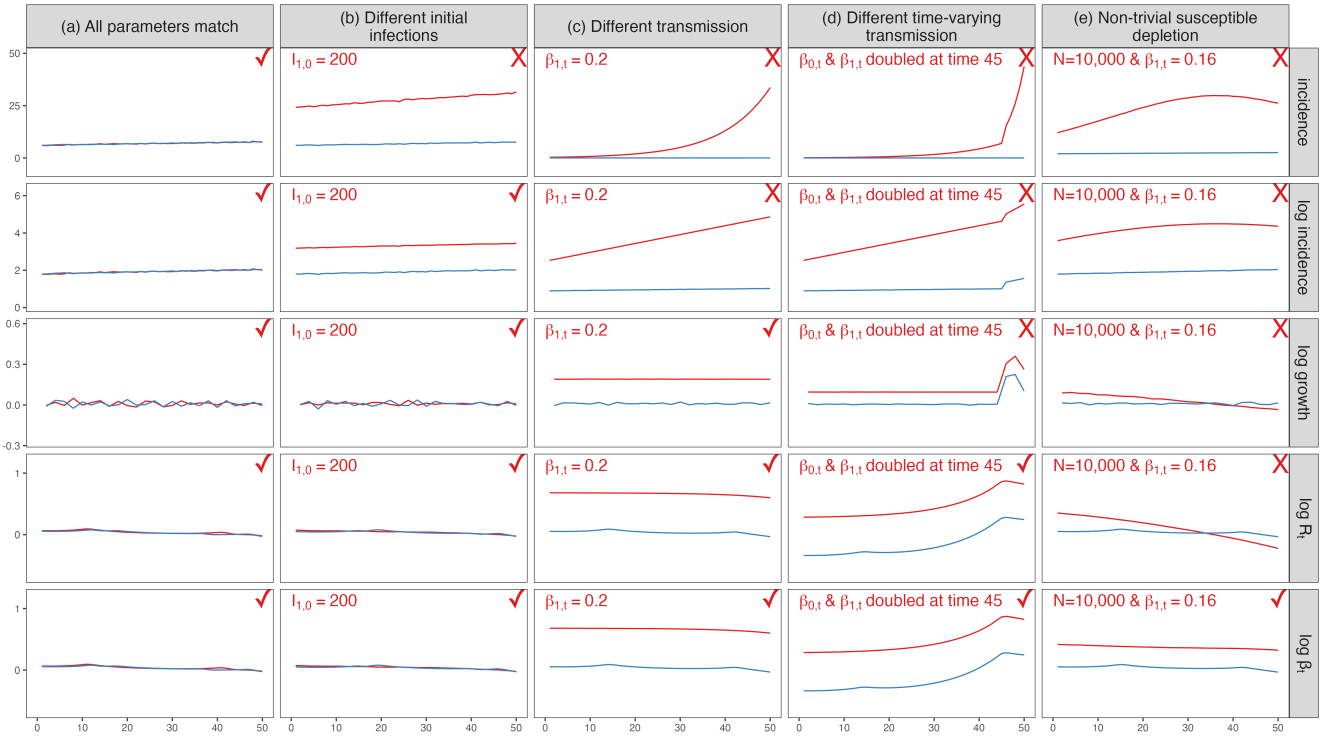

Fig. 1. Comparison of different model specifications

We vary the choice of outcome variable across rows and the underlying data-generating process across columns. We simulate data from an SIR model and average outcomes over 1000 draws. Lines indicate the average outcome value for two units. Assume unit 0 is graphed in blue, and unit 1 is graphed in red. There are no treatment effects in any scenario; therefore plots display untreated potential outcomes. If trends are parallel (checkmark), then DiD would estimate unbiased treatment effects with the corresponding model specification. Baseline parameters are $N = 10^8$, $I_0 = 50$, $\frac{1}{\gamma} = 10$, $\beta_t = 0.105 \ \forall t$. (a) "All parameters match" matches all parameters between units for all t. (b) "Different initial infections" changes the initial number of infections for the unit graphed in red to $I_{1,0} = 100$. (c) "Different transmission" sets $\beta_{1,t} = 0.2$. (d) "Different time-varying transmission" sets $\beta_{1,t} = 0.2$ for the treated unit and doubles both $\beta_{0,t}$ and $\beta_{1,t}$ for time $t \geq 45$. (e) "Non-trivial susceptible depletion" sets a smaller population size $N = 10^4$ for both units and sets $\beta_{1,t} = 0.160$.

The implications of this condition are less straightforward than those of prior specifications. First, the log growth specification again imposes no requirements on expected incidence at t_1. Furthermore, although these results invoke the "infinite susceptible population" assumption, this is less restrictive in the log growth specification, a result of dividing incidence at adjacent time-steps.

In the special case where the length of generation interval is equal to the time step ($\gamma = 1$, i.e., no intergenerational compounding)[‡], Proposition 5 becomes:

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

Effective contact rates could then be allowed to differ between groups and vary over time, providing that their ratios remain constant across groups. In practice, this means that unbiased treatment effect estimation could be achieved when units differ in terms of baseline mask mandates or vaccination coverage, provided this ratio remains unchanged over the study period. However, this condition imposes restrictions on both the generation interval (γ) and susceptible depletion (Figure 1, panels 3(c) through 3(e)).

Proposed model specifications. Of the specifications characterized above, log growth is the most flexible, allowing the effective contact rate to differ across groups by a constant ratio under certain conditions. However, these may be too narrow for many applications. We therefore propose two alternatives that allow us to relax assumptions further by drawing on common epidemiological quantities: the log of the effective reproduction number (R_t) and of the effective contact rate (β_t).

Log effective reproduction rate **(R_t).** The effective reproduction number, traditionally denoted R_t,[§] measures the average number of secondary infections caused by each individual infection over a time interval traditionally denoted t. This quantity is commonly used to understand the risk of exponential spread in a population (19), accounting for contact patterns, precautionary behaviors, and immunity (20). It is often benchmarked against a value of 1: if $R_t < 1$, incident infections will decrease over time, and the goal of policies is often to maintain R_t below 1 (18, 19, 21).

In this section, we consider modeling $\log(R_t)$ as the outcome variable in DiD and demonstrate that this model specification further relaxes the assumptions about the underlying transmission dynamics. Although there are a few possible formulations of the effective reproduction number based on β_t (22), we use the popular cohort-based definition (21), which can be more easily mapped to intervention timing than "instantaneous" approaches (e.g., Cori et al. (19)).

[‡]"No intergenerational compounding" implies that individuals who become infectious at time t recover by the next time step $t+1$.

[§]Because R is traditionally used both to denote the "Recovered" individuals and the effective reproduction number, we follow this convention. In text that follows, R refers always to the effective reproduction number.

Proposition 6 (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:*

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

.

With this link (proof in Appendix B), we can characterize the conditions required for the parallel trends assumption to hold when the outcome specification is $\log R_{d,t}$.

Proposition 7 (Parallel trends: Log R_t (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1), log-transformed effective reproduction number model specification ($Y_{d,t} = \log(R_{d,t})$, $g(\cdot) = \log(\cdot)$), the "infinite susceptible population" parallel trends assumption (Eq. 3) holds if and only if*

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

Per Proposition 7, the log R_t specification requires only a constant ratio between effective contact rates in treatment and comparison groups during time-steps included in the R_t estimate. With a short generation interval, aggregation presents few practical concerns; with longer generational intervals, researchers may be wary about spillover between treatment and comparison timings.

Like other specifications discussed, this specification also is sensitive to differential susceptible depletion. For small populations or infectious diseases with intense transmission, the susceptible population may deplete quickly and may do so differentially across treated and comparison groups if they start with different susceptible fractions. In this case, DiD using log R_t as an outcome would produce biased treatment effect estimates (Figure 1, panel 4(e)). In the context of COVID-19, population susceptibility often declined rapidly during major waves, underscoring the value of having an model specification robust to non-trivial susceptible depletion.

Log effective contact rate **(β_t).** Although the effective reproduction number is more often modeled in applied epidemiology, the effective contact rate may be a more robust outcome. Proposition 8 formalizes the conditions required for the parallel trends assumption to hold when using a log β_t model specification.

Proposition 8 (Parallel trends: Log β_t)**.** *Assuming an SIR data-generating process (Eq. 1) and a log-transformed effective reproduction number specification ($Y_{d,t} = \log(\beta_{d,t}$, $g(\cdot) = \log(\cdot)$) the parallel trends assumption (Eq. 3) holds if and only if*

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

Proposition 8 follows from substituting $Y_{d,t} = \beta_{d,t}$ into the log-transformed parallel trends assumption as defined in Eq. 3. This specification simplifies the assumptions imposed by R_t, requiring only a constant ratio on β_t and is not sensitive to susceptible depletion.

Estimation.

R_t and β_t. Our two proposed outcomes, R_t and β_t, are not straightforward transformations of incident infections, but can be obtained via maximum likelihood estimation assuming an SIR model. For example, under the SIR data-generating process (Eq. 1), we can obtain a maximum likelihood estimate of β_t:

Proposition 9 (Estimation of $\beta_{d,t}$)**.** *Assuming an SIR data-generating process (Eq. 1), 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:*

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

Beyond SIR models, we can also estimate R_t and β_t assuming more complex data-generating processes, another benefit over incidence, log incidence, or log growth specifications discussed further in **Extensions**.

ATTs. We can estimate ATTs of interest through regression:

$$g\left(\mathbb{E}[Y_{d,t}]\right) = \alpha + \eta_d + \tau_t + \delta\left(\text{TreatedUnit}_d \times \text{PostTreatment}_t\right), \quad [4]$$

where $Y_{d,t}$ refers to the chosen model specification, α denotes the intercept, η_d denotes the unit fixed-effect for each unit d, and τ_t denotes the time fixed effect for each time period t. TreatUnit_d and PostTreatment_t are binary variables indicating whether unit d is ever treated and whether the time period t is post-intervention, respectively. With this setup, δ is the ATT of interest. When $g(.)$ is the identity, we use OLS, and for a log link, we use Poisson regression. (As mentioned above, we could often nearly equivalently model $\mathbb{E}\left(\log[Y_{d,t}]\right)$ with OLS, but Poisson regression circumvents the issues of managing zeros in the outcome variable with a log transformation (10)). We use the wild score bootstrap for inference (Appendix G) (23, 24).

Average marginal effects. When using non-incidence model specifications, estimated regression coefficients may not provide useful interpretations beyond their signs. To improve interpretability, we convert coefficients from non-incidence specifications back to the incidence scale by calculating their average marginal effects (AME), enabling comparison across different model specifications.

For log incidence and log growth models, we first estimate untreated potential outcomes for the treated group using the observed outcome trajectories in the treated units and the estimated ATT and convert each to the incidence scale. We then compute the difference between potential and observed outcomes in the post-intervention period, and average the differences over all units to obtain the AME. For log R_t and log β_t models, we estimate the potential outcomes for each unit by simulating infections from an infectious disease transmission model. We simulate the treated and untreated potential trajectories for treated units 1000 times and calculate the difference in the post-intervention period to estimate AMEs. Detailed algorithms are provided in Appendix E.

Extensions.

Generalizations of SIR models. In practice, epidemiologists rarely assume simple SIR data-generating processes, in favor of models that capture more granular transmission dynamics. For example, many models introduce an additional exposed state (Susceptible-Exposed-Infectious-Recovered [SEIR]), which tracks exposed but not yet infected individuals (Appendix F) (25). Models can also be adapted to include heterogeneity across subgroups or to agent-based frameworks that allow heterogeneity in individual agents.

Our proposed model specifications involving the effective reproduction number and the effective contact rates remain conceptually meaningful quantities and can still be estimated assuming more complex transmission dynamics. Wallinga and Teunis's cohort-based R_t estimator can accommodate a wide range of underlying data-generating processes, producing an unbiased estimate when a directed network for infectious disease transmission is defined and the generation interval is known or can be estimated (21). When the time-step is approximately equal to the generation interval, estimates of R_t can be converted to β_t by dividing by the susceptible population fraction. (For COVID-19 applications detailed below, we assume a generation time of approximately a week and model R_t at weekly intervals). More complex approximation of a time-varying β_t, rather than assuming constant β_t over a generation interval, may be needed when generation intervals are longer.

Time-step aggregation. When a disease is relatively rare, researchers may choose to aggregate data over time above the level of the infectious disease process timestep (e.g., aggregating days into weeks or months). In this case, we require the aggregated parallel trends assumption:

$$g\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_1\right]\right) - g\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_0\right]\right) = g\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_1\right]\right) - g\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_0\right]\right), \quad [5]$$

where $\mathcal{T}_0$ and $\mathcal{T}_1$ are the collections of pre- and post-treatment time periods. Here for Eq. 5 with an identity or a log link to be true, it is sufficient to assume that the original parallel trends assumption (Eq. 3) holds for every two time periods t_1 and t_2 such that $1 \leq t_1 \leq T_0 < t_2 \leq T$, where T_0 is the time of intervention. This result follows from the linearity of expectation for the identity link; for a log link, we derive the condition in Supplement Proposition 2.

Multiple units and time periods. In practice, researchers often have more than 2 periods and more than 2 units in their study population. We can extend the parallel trends assumption (Eq. 3) to both multiple units and multiple time periods:

$$g\left(\mathbb{E}\left[Y_{d,t}(0)\middle|d \in \mathcal{N}_1, t \in \mathcal{T}_1\right]\right) - g\left(\mathbb{E}\left[Y_{d,t}(0)\middle|d \in \mathcal{N}_1, t \in \mathcal{T}_0\right]\right) = g\left(\mathbb{E}\left[Y_{d,t}(0)\middle|d \in \mathcal{N}_0, t \in \mathcal{T}_1\right]\right) - g\left(\mathbb{E}\left[Y_{d,t}(0)\middle|d \in \mathcal{N}_0, t \in \mathcal{T}_0\right]\right), \quad [6]$$

where $\mathcal{N}_1$ and $\mathcal{N}_0$ denote the collections of treated and control units and $\mathcal{T}_0$ and $\mathcal{T}_1$ are the collections of pre- and post-treatment time periods. In this context, we can estimation procedures following those outlined in **Estimation**.

Multiple pre- and post-time periods With multiple pre- and post-time periods, we can meet the conditions for Eq. 6 with the same parallel expected trajectories condition required for time-step aggregation (Corollary 2). This translates to similar conditions required for each model specification in Propositions 3-5 and 7-8, except that conditions must hold for all time period pairs, rather than just t_1 and t_2.

In the special case of staggered intervention roll-out, simply averaging over the post-intervention treatment effects does not yield a sensible estimand unless additional homogeneity assumptions are imposed (14). Recently developed estimators allow for more flexible treatment effect estimation, such as those proposed by Callaway and Sant'Anna (26), Sun and Abraham (27) and Borusyak et al. (28). In general, these methods proposed ways to construct a group-time treatment effect for each treatment-adoption cohort, aggregating these group-time effects to obtain an overall ATT, and the above assumption would again suffice for unbiased effect estimation.

Multiple treated and comparison units With multiple treated units, Eq. 6 requires that transformed average differences between pre- and post-treatment outcomes be equal across treated and untreated units. This too follows from assuming parallel trends between pairs of treated and comparison units. Outside of this scenario (or scenarios with equal data-generating parameters for all units within treatment or comparison groups), researchers may need to exercise care in identifying appropriate groups, particularly for incidence specifications, as equal averages across data-generating parameters for treatment and comparison groups is not sufficient for equal expected incidence (Proposition 2).

Power Analysis

Power to detect treatment effects may vary across model specifications (29). We therefore performed a simulation study to explore the performance of the five model specifications in the previous section.

Simulation setup. With an SIR model (Eq. 1), we simulated infection trajectories in 50 units over a total time interval of 17 weeks, with an average generation interval of 10 days $\left(\gamma = \frac{1}{10}\right)$. Twenty-five units (50% of all units) were randomly designated as treated units. We introduced 100 seed infections to the population at time 0 and set a constant baseline effective contact rate $\beta_t \in \{0.100, 0.115\}$. We then varied the baseline effective contact rate in treated units to be either equal to that of untreated units or 10% greater. For treated units, we introduced an intervention that changed the effective contact rate at week 9. We varied the magnitude of the treatment effect as a multiplicative factor (δ) on the baseline transmission rate, by setting $0.70 \leq \delta \leq 1.3$. The scenario $\delta = 1$ corresponded to the null-effect case, i.e., no change in transmission is introduced in the treatment group. We also varied the population size in each unit, $N \in \{5000, 10000\}$. Simulation specifications are also summarized in Table A1.

We estimated the five outcomes outlined in Table 1. For R_t, we used the Wallinga-Teunis estimator and divided estimates by the susceptible fraction to estimate β_t. After estimating these outcome, we excluded the first 5 weeks of data for all specifications because previous studies have noted R_t estimates are often biased at the start of a time series (22).[¶] We also discarded the last 5 weeks of data to account for a similar bias due to delayed reporting (21).

We estimated treatment effects per **Estimation** and reported the probability of rejecting the null in each scenario using confidence intervals generated by wild score bootstrap (24). These corresponded to power in the case where $\delta \neq 1$ and type I error where $\delta = 1$. We used a significance level of 0.05, and all results shown are an average of 1000 simulations.

Simulation results. Figure 2 summarizes the power ($\delta \neq 1$) and type I error ($\delta = 1$) when initial infections, initial susceptible fractions, and transmission parameters are all equal between the treatment and control groups. The left and right panels differ in the scale of effect size: the left plots the effect size (δ) which is a multiplicative factor on the baseline effective contact rate, and the right plots the percentage change in infection due to the change in effective contact rate.

In this case, all specifications controlled type I error (horizontal dashed line at 0.05) absent a treatment effect. With a non-null treatment effect, incidence and log incidence specifications produced the highest power, with minimal difference between the two. Of the remaining specifications, the log growth specification had the lowest power, while log R_t and log β_t had greater power, albeit slightly less than incidence and log incidence specifications. For example, with a treatment effect corresponding to a 5% increase in cumulative incidence over 6 weeks (e.g., 500 additional infections per a population of 10,000 individuals), the log growth specification had 80% power, while the other four model specifications had power exceeding 95%.

When the baseline effective contact rate differs between treatment and comparison groups, all methods except log β_t produced biased treatment effect estimates, failing to control type I error at $\delta = 1$ (Figure 3). Although the fraction of the susceptible population depleted was lower with $N = 10,000$, type I error remained higher than in the $N = 5,000$ due to higher power with a larger population.

Case Studies

Removing school mask mandates in Massachusetts.

Background. We last explore the implications of different specifications through two published case studies. We first consider an analysis of school mask mandates. On February 28, 2022, Massachusetts lifted its universal mask requirement in schools. Seventy of the 72 school districts lifted mask mandates shortly thereafter, but two districts, Boston and Chelsea, did not remove mask requirements until June 2022. The *New England Journal of Medicine* published a staggered DiD analysis estimating the impact of district-level mask policies, with weekly COVID-19 cases per 1000 as the outcome variable, using data from September 2021 to June 2022 (5). The analysis showed a statistically and practically significant

[¶] There are some other approaches established in the literature to adjust when analyzing empirical data, such as inferring the unobserved initial generations of infections using a linear exponential growth model (30).

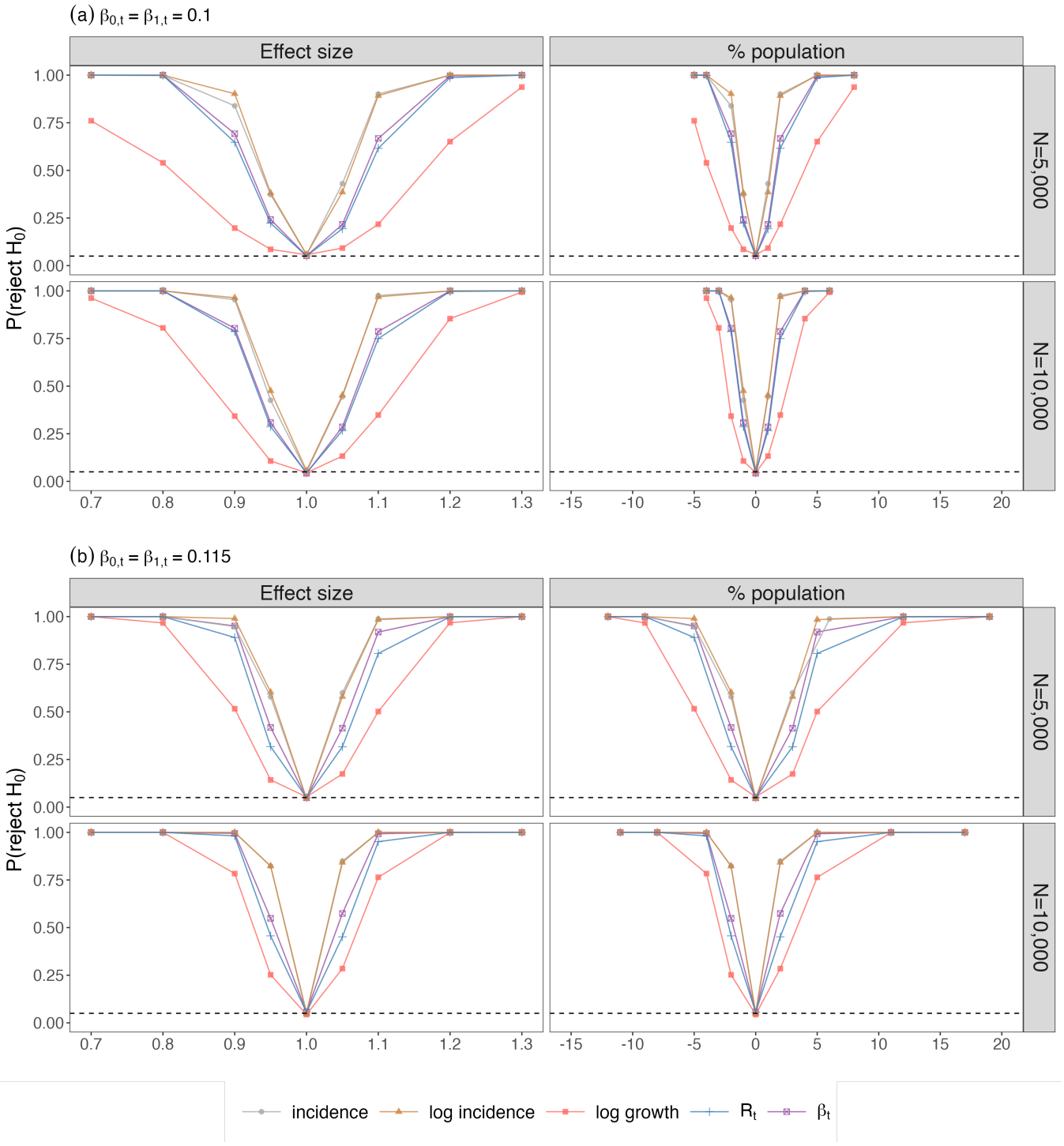

Fig. 2. Power of models with all parameters equal between treatment and comparison groups

The top and bottom plots differ in the baseline effective contact rates: (a) $\beta_{0,t} = \beta_{1,t} = 0.1$, and (b) $\beta_{0,t} = \beta_{1,t} = 0.115$. Within each plot, left and right panels differ in the scale of effect size: the left plots the effect size as a multiplicative effect on baseline $\beta_{d,t}$, and the right the percentage change in cumulative incidence. Across rows, we vary the population size. We generated data from an SIR model per Eq. 1. Solid colored lines indicate the average probability of rejecting the null over 1000 simulations. The dashed line indicates a significance level of 0.05.

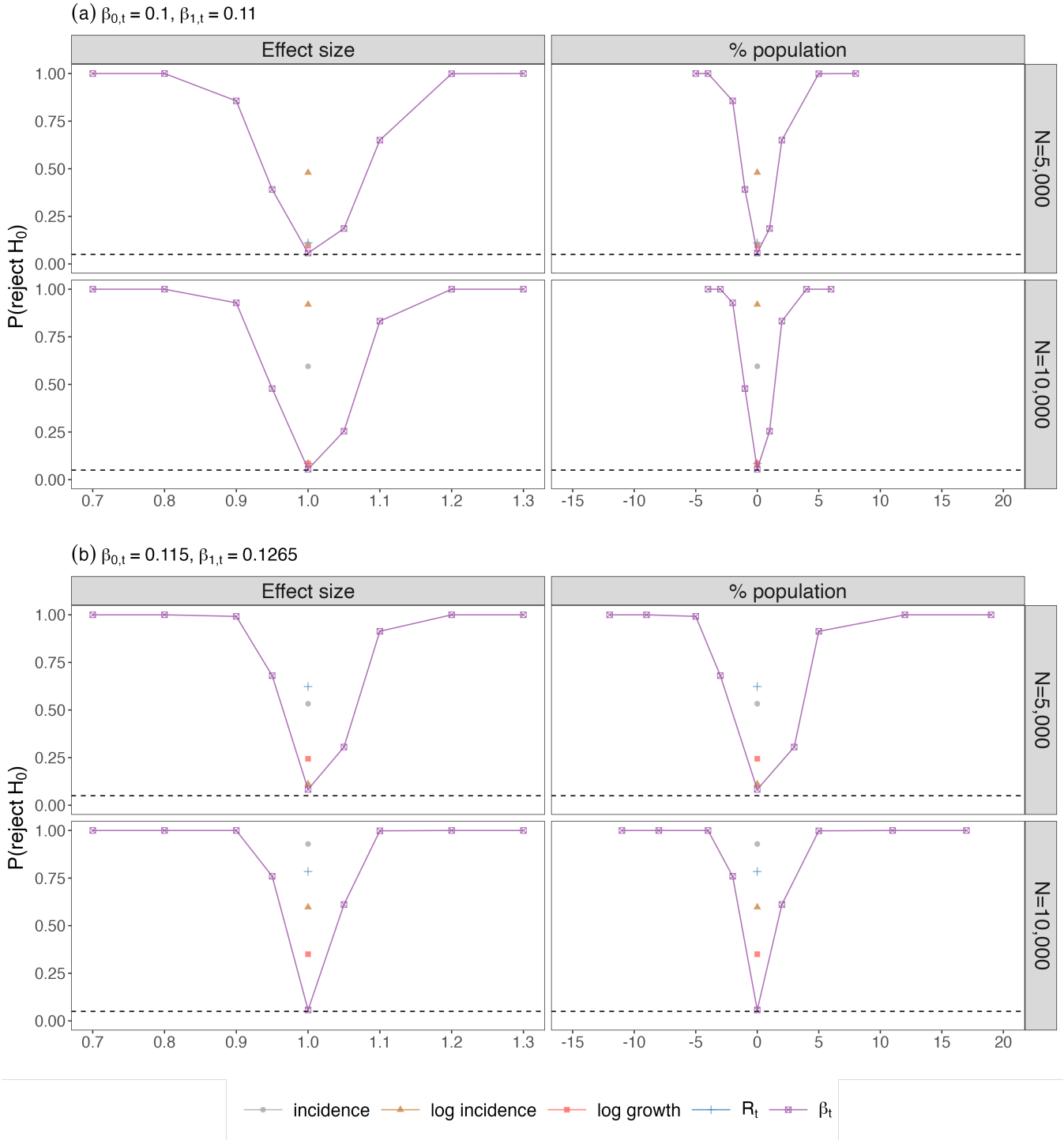

Fig. 3. Power of models with baseline effective contact rate in treated units to be 10% greater than that of untreated units

The top and bottom plots differ in the scale of effect size: (a) $\beta_{0,t} = 0.1$, $\beta_{1,t} = 0.11$, and (b) $\beta_{0,t} = 0.115$, $\beta_{1,t} = 0.1265$. The left and right panels differ in the scale of effect size: the left plots the effect size as a multiplicative effect on baseline $\beta_{d,t}$, and the right the percentage change in cumulative incidence. Across rows, we vary the population size. We generated data from an SIR model per Eq. 1. The solid colored line indicates the average probability of rejecting the null over 1000 simulations in the β_t model. Points at $\delta = 1$ indicate the failure to control type I errors in all other models. The dashed line indicates a significance level of 0.05.

increase in cases (an additional 45 cases per 1000 students and staff) associated with the removal of school mask mandates.

We extended this analysis with log incidence and log growth specifications. We did not estimate effective reproduction number and effective contact rate estimates because the these are not typically estimated for subpopulations or below the county-level (22, 31).

Re-analysis methods. We applied the authors' DiD approach to incidence, log incidence, and log growth models, which is similar to the specification in **Estimation**, but accounts for staggered treatment roll-out by estimating and averaging group-time average treatment effects ($ATT(g,t)$), where groups are defined based on time of treatment. We first replicated their specification using the Callaway & Sant'Anna estimator (26), with results presented in Supplement Table A2. As log incidence and log growth specifications do not allow zero outcomes, we implemented these with Poisson regression, first using the Callaway & Sant'Anna estimator and then a similar alternative by Sun and Abraham (27), that more straightforwardly supports inference with Poisson regression. For both, we present the average over all $ATT(g,t)$. We then calculated AMEs for log incidence and log growth treatment effects.

Re-analysis results. Figure A1 plots district-level COVID-19 cases per 1000 over the study period. Using the Callaway & Sant'Anna estimator, we replicated the authors' original findings, with an estimated 46 additional cases per 1000 students and staff (Figure A2). With the Sun & Abraham estimator, there was similarly an increase of 48 cases per 1000 students and staff (95% CI: 40, 57) (Table 2).

Figure A1 suggests that clusters do not have the same initial levels. Over the full study period, estimated treatment effects declined with a log incidence specification to 13 (95% CI: -31, 47) cases per 1000 students and staff (Table 2). The point estimate in the log growth specification was near 0, but highly uncertain; we could not rule out differences up to 120 cases per 1000 students and staff. However, with a shorter post-intervention horizon, both the incidence and the log incidence specifications detected a statistically and practically significant increase of 7-9 cases per 1000 students and staff over 5 weeks following the lifting of school mask mandates, with an effect of similar magnitude (but more uncertainty) in the log growth specification.

Kansas mask mandates.

Background. On July 3, 2020, Kansas adopted an executive order requiring face masks in most indoor and outdoor public places. As this was initially only adopted by 15 counties (32), *JAMA Network Open* published a DiD analysis evaluating the effect of county-level mask mandate adoption on daily COVID-19 rates (33). The authors used incidence specifications, where the 7-day moving average of COVID-19 cases per 100,000 was used as the outcome variable, and the policy was evaluated at 21 days after the executive order was effective (i.e., July 24, 2020 was the intervention time) to account gradual uptake and the lagged impact of behavior on cases. The authors found a statistically and practically significant reduction in COVID-19 cases (-20 cases per 100,000 with 95% CI: -27 to -14) following the executive order.

Re-analysis methods. We first replicated the authors' original specification with small modifications:

$$Y_{d,t} = \alpha + \eta_d + \tau_t + \\ \delta\left(\text{TreatedCounty}_d \times \text{PostTreatment}_t\right) + \epsilon_{d,t},$$

where the outcome $Y_{d,t}$ was the county-level number of cases per 100,000 population (i.e., an incidence model), and the remaining variables and coefficients were defined as in **Estimation**. In our re-analysis, we aggregated data to the week-level and excluded counties with zero reported COVID-19 cases prior to the announcement of mask mandates. We also excluded covariates (an indicator for no reported cases and the number of days since the first recorded case in the county) to avoid recent methodological concerns about verifying the parallel trends assumption conditional on covariates (34).

We then re-analyzed the Kansas mask mandate example with the other four specifications. For log R_t and log β_t models, we obtained county-level data on R_t from COVIDestim nowcasting (31) and translated this to β_t by dividing R_t by the estimated susceptible fraction as illustrated in Proposition 6. We used estimated the susceptible fraction using cumulative infections from COVIDestim nowcasting (31) and each county's 2019 population data from the Census Bureau (35) to impute susceptible fractions. For AMEs with $\log(R_t)$ and $\log(\beta_t)$, we used an SEIR model, parameterizing for COVID-19, with an average of 5 days of infectiousness and a 3-day incubation period (36).

Re-analysis results. Our treatment effect estimate using a slightly modified incidence model was -22 cases per 100,000 (95% CI: -28 to -15), very similar to the original estimate (-20 cases per 100,000 with 95% CI: -27 to -14).

Table 2 summarizes the estimated treatment effects and average marginal effects obtained using all five specifications. We observed a similar treatment effect in the log incidence specification compared to the incidence specification. Effects were smaller and insignificant for log growth and log R_t specifications. However, in this application, many counties experienced considerable susceptible depletion (e.g., the estimated susceptible fraction dropped by 40% in Nemaha and 43% in Norton over the study period), a potential threat to the validity of log growth and log R_t models. By contrast, the log β_t specification produced a significant treatment effect at 90% confidence level, with an average marginal effect of -11 cases per 100,000.

Discussion

DiD is a popular method for observational causal inference in health policy. This arises in part from its flexibility: DiD allows researchers to estimate treatment effects, even absent comparison groups that exactly match the treatment group. DiD is traditionally agnostic as to why treatment and comparison groups differ in level but are believed to have matching trends (11). However, by synthesizing DiD with a mechanistic model for infectious disease transmission, we showed how context-specific theory informs interpretation of the parallel trends assumption.

Assuming an underlying SIR framework, our paper first formalized the epidemiological conditions under which the parallel trends assumption holds for the three common model

Table 2. Results from re-analyzing the effect of face mask policies in empirical examples

Study	Length of post-intervention periods	Outcome specification	Treatment effect[1] (95% CI)	Average marginal effect[2] (95% CI)
Removing school mask mandates in Massachusetts [4]	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)
	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)
Mask mandates in Kansas counties [5]	19 weeks	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† (0.86, 1.01)	−10.8† (−36.0, 2.6)

1 Treatment effect refers to the point estimate in coefficients obtained from OLS (incidence model) or average relative risk per unit from Poisson regression (other four models)
2 Average treatment effects are imputed as described in Section **Average marginal effects** and Appendix E. AMEs for $\log R_t$ and $\log \beta_t$ models are calculated assuming an SEIR transmission process per Section F
* represents significance at $\alpha = 0.05$
† represents significance at $\alpha = 0.10$

specifications in the COVID-19 literature: incidence, log incidence, and log growth. Although modeling incidence was popular in the COVID-19 medical and health policy literature (Appendix A), this approach requires identical data-generating processes between all treated and comparison units, as others have previously noted (4). Modeling log incidence allows for different numbers of initial infections but still requires equal effective contact rates and an "infinite susceptible population" assumption. Nevertheless, because we found incidence and log incidence specifications to have very similar power, we recommend that researchers default to the latter, binning data or applying Poisson regression if there are zero-valued outcomes (37). By contrast, although log growth specifications allow for more flexibility, they comes with a significant power cost.

We argued that the log growth specification approximates assuming parallel trends in the effective contact rate (e.g., assuming similar behavior but different vaccination rates across units). Both modeling log growth and the log of effective reproduction number (a similar but higher power alternative) may produce biased estimates in the context of susceptible depletion. Our most robust alternative is therefore modeling the log of effective contact rate (β_t) directly. We showed that this approach has higher power than modeling log growth, performs well under more flexible assumptions, and can be estimated with zero-valued outcomes using Poisson regression. We also demonstrated that this can be applied to more diverse and complex underlying transmission processes. As a result, our work provides a useful bridge between post-hoc policy evaluation and prospective transmission dynamic models that serve as a backbone of infectious disease epidemiology.

Last, we apply different model specifications to previously published examples, highlighting trade-offs between bias and power and emphasizing the importance of evaluating and discussing the plausibility of assumptions required for a chosen specification.

This work has several limitations. First, even our most robust specifications may not be plausible in practice. Nevertheless, our approach allows for more flexibility than many proposed alternatives. For example, Callaway and Li (4) proposed an alternative estimator in the context of early COVID-19 that assumed unconfoundedness after conditioning on all pre-treatment period outcomes. This method performs well in the early stages of an epidemic when conditions across units in both groups are similar. However, it may be challenging to find exact matching units over the full epidemic history, especially when the effective contact rate is time-varying due to shifting precautionary behaviors or vaccination rates. Relatedly, although our approach can be extended to a wide array of models, we do not consider the implications of spillovers, which may obscure treatment effects.

Second, estimating the effective reproduction number and the effective contact rate can be challenging in practice. Many past studies (e.g., O'Driscoll et al. (20), Gostic (22)) have discussed challenges and proposed approaches for translating observed cases and deaths into estimates of infections and the effective reproduction number, which would be valuable when applying our methods. Last, our results highlight that in many cases, policy evaluations may be underpowered to detect effects of substantive interest. Future efforts could take a decision-analytic approach to integrate the results of policy evaluations with costs and benefits of implementing different model specifications.

Overall, our work provides a framework for integrating transmission dynamics into DiD and proposes novel methods to support rigorous evaluation of infectious disease policies.

ACKNOWLEDGMENTS. The authors gratefully acknowledge feedback from Tori Cowger, Jeremy Goldwasser, Kathryn T. Hall, Joseph Hogan, Youjin Lee, Eleanor Murray, Jonathan Roth, and Pedro Sant'Anna. We also greatly appreciate replication code shared by Donna Ginther. This work was supported by the Centers

for Disease Control and Prevention through the Council of State and Territorial Epidemiologists (NU38OT000297-02).

1. Coady Wing, Kosali Simon, and Ricardo A Bello-Gomez. Designing difference in difference studies: best practices for public health policy research. *Annual review of public health*, 39: 453–469, 2018.
2. Charles Courtemanche, Joseph Garuccio, Anh Le, Joshua Pinkston, and Aaron Yelowitz. Strong social distancing measures in the united states reduced the covid-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed covid-19 cases across the united states. *Health affairs*, 39(7):1237–1246, 2020.
3. Victor Chernozhukov, Hiroyuki Kasahara, and Paul Shrimpf. The association of opening k-12 schools and colleges with the spread of covid-19 in the united states: country-level panel data analysis. 2021.
4. Brantly Callaway and Tong Li. Policy evaluation during a pandemic. *Journal of Econometrics*, 236(1):105454, 2023.
5. Tori L Cowger, Eleanor J Murray, Jaylen Clarke, Mary T Bassett, Bisola O Ojikutu, Sarimer M Sánchez, Natalia Linos, and Kathryn T Hall. Lifting universal masking in schools—covid-19 incidence among students and staff. *New England Journal of Medicine*, 387(21):1935–1946, 2022.
6. Timo Mitze, Reinhold Kosfeld, Johannes Rode, and Klaus Wälde. Face masks considerably reduce covid-19 cases in germany. *Proceedings of the National Academy of Sciences*, 117 (51):32293–32301, 2020.
7. Thiemo Fetzer and Thomas Graeber. Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. *Proceedings of the National Academy of Sciences*, 118(33):e2100814118, 2021.
8. Joshua A Salomon. Integrating economic evaluation and implementation science to advance the global hiv response. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 82:S314–S321, 2019.
9. Wan Yang and Jeffrey Shaman. Reconciling the efficacy and effectiveness of masking on epidemic outcomes. *medRxiv*, pages 2023–05, 2023.
10. Jonathan Roth and Pedro HC Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747, 2023.
11. Ariella Kahn-Lang and Kevin Lang. The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3):613–620, 2020.
12. Noah A Haber, Emma Clarke-Deelder, Avi Feller, Emily R Smith, Joshua A Salomon, Benjamin MacCormack-Gelles, Elizabeth M Stone, Clara Bolster-Foucault, Jamie R Daw, Laura Anne Hatfield, et al. Problems with evidence assessment in covid-19 health policy impact evaluation: a systematic review of study design and evidence strength. *BMJ open*, 12(1):e053820, 2022.
13. Emanuele Amodio, Michele Battisti, Antonio Francesco Gravina, Andrea Mario Lavezzi, and Giuseppe Maggio. School-age vaccination, school openings and covid-19 diffusion. *Health Economics*, 32(5):1084–1100, 2023.
14. Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.
15. Jeffrey M Wooldridge. Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26(3):C31–C66, September 2023. ISSN 1368-4221. . URL https://doi.org/10.1093/ectj/utad016.
16. Emilia Vynnycky and Richard White. *An Introduction to Infectious Disease Modelling*. Oxford University Press, Oxford ; New York, 1st edition edition, July 2010. ISBN 978-0-19-850808-3.
17. Norman TJ Bailey et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
18. Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.
19. Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 2013.
20. Megan O'Driscoll, Carole Harry, Christl A Donnelly, Anne Cori, and Ilaria Dorigatti. A comparative analysis of statistical methods to estimate the reproduction number in emerging epidemics, with implications for the current coronavirus disease 2019 (covid-19) pandemic. *Clinical Infectious Diseases*, 73(1):e215–e223, 2021.
21. Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6):509–516, 2004.
22. et al. Gostic, Katelyn M. Practical considerations for measuring the effective reproductive number, rt. *PLoS computational biology*, 16(12):e1008409, 2020.
23. A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, 90(3):414–427, 2008.
24. James G MacKinnon, Morten Ørregaard Nielsen, and CREATES Matthew D Webb. Fast and wild: Bootstrap inference in stata using boottest. *Queen's University Economics Department working paper*, (1406), 2018.
25. Anas Abou-Ismail. Compartmental models of the covid-19 pandemic for physicians and physician-scientists. *SN comprehensive clinical medicine*, 2(7):852–858, 2020.
26. Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.
27. Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.
28. Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2021.
29. Alyssa Bilinski and Laura A Hatfield. Nothing to see here? non-inferiority approaches to parallel trends and other model assumptions. *arXiv preprint arXiv:1805.03273*, 2018.
30. Andrea Brizzi, Megan O'Driscoll, and Ilaria Dorigatti. Refining reproduction number estimates to account for unobserved generations of infection in emerging epidemics. *Clinical Infectious Diseases*, 75(1):e114–e121, 2022.
31. Marcus Russi and Melanie Chitwood. *covidestim: Real-time Bayesian forcasting of R_t*, 2022.
32. Kevin Dayaratna and Norbert Michel. A statistical analysis of mandates and mask usage in kansas. Technical report, Heritage Foundation Special Report, 2021.
33. Donna K Ginther and Carlos Zambrana. Association of mask mandates and covid-19 case rates, hospitalizations, and deaths in kansas. *JAMA network open*, 4(6): e2114514–e2114514, 2021.
34. Carolina Caetano, Brantly Callaway, Stroud Payne, and Hugo Sant'Anna Rodrigues. Difference in differences with time-varying covariates. *arXiv preprint arXiv:2202.02903*, 2022.
35. United States Census Bureau. County population totals: 2010-2019, 2019. https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/cities/totals/sub-est2019_20.csv.
36. Alyssa Bilinski, Andrea Ciaranello, Meagan C Fitzpatrick, John Giardina, Maunank Shah, Joshua A Salomon, and Emily A Kendall. Estimated transmission outcomes and costs of sars-cov-2 diagnostic testing, screening, and surveillance strategies among a simulated population of primary school students. *JAMA pediatrics*, 176(7):679–689, 2022.
37. Jiafeng Chen and Jonathan Roth. Logs with Zeros? Some Problems and Solutions*. *The Quarterly Journal of Economics*, page qjad054, December 2023. ISSN 0033-5533. . URL https://doi.org/10.1093/qje/qjad054.
38. Yee Teh, David Newman, and Max Welling. A collapsed variational bayesian inference algorithm for latent dirichlet allocation. *Advances in neural information processing systems*, 19, 2006.

A. Literature Review

Year	Journal	Authors	Title	Main Outcomes	Model Type	Intervention	Country	Justification for Outcomes	Sensitivity Analysis
2020	Health Affairs	Courtemanche, et al.	Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate	Natural log of cumulative daily COVID-19 cases minus that on the prior day	Log incidence	Multiple Interventions	USA		• Different Timing of Interventions • Outliers • Adding/Removing Variables • Alternative Functional Forms
2020	Health Affairs	Lyu & Wehby	Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US	Natural log of cumulative COVID-19 cases on a given day minus that in the prior day	Log incidence	Mask Mandate	USA		• Different Timing of Interventions • Adding/Removing Variables
2020	Health Affairs	Lyu & Wehby	Shelter-In-Place Orders Reduced COVID-19 Mortality And Reduced The Rate Of Growth In Hospitalizations	Natural log of cumulative cases on a day minus that on the prior day; Inverse hyperbolic sine transformation	Log incidence	Shelter-in-place	USA	An inverse hyperbolic sine transformation handled zeroes and allowed for a similar interpretation of the daily growth rate	
2020	Health Affairs	Pichler, Wen, & Ziebarth	COVID-19 Emergency Sick Leave Has Helped Flatten The Curve In The United States	New state-level cases	Incidence	Paid Sick Leave	USA		
2020	JAMA Network Open	Lyu & Wehby	Comparison of Estimated Rates of Coronavirus Disease 2019 (COVID-19) in Border Counties in Iowa Without a Stay-at-Home Order and Border Counties in Illinois With a Stay-at-Home Order	Cumulative cases of COVID-19 per 10,000	Incidence	Stay-at-home Order	USA		• Different Timing of Interventions • Adding/Removing Variables
2020	Nature	Hsiang, et al.	The effect of large-scale anti-contagion policies on the COVID-19 pandemic	Differences of natural logarithm as per-day growth rate of infections	Log incidence	Multiple Interventions	Multiple Countries	Motivated with an SIR model of disease contagion	• Different Timing of Interventions • Outliers • Adding/Removing Variables • Lagged Outcome or Covariates
2020	PNAS	Mitze, et al.	Face masks considerably reduce COVID-19 cases in Germany	Number of cumulative COVID-19 cases per 100,000	Incidence	Mask Mandate	Germany		• Different Timing of Interventions • Adding/Removing Variables • Alternative Functional Forms • Alternative Estimation Models
2020	PNAS	Taylor, et al.	Livestock plants and COVID-19 transmission	COVID-19 weekly log difference	Log incidence	Livestock Plant Shutdowns	USA		• Outliers • Adding/Removing Variables • Alternative Functional Forms • Alternative Estimation Models • Alternative Variable Definitions
2021	Health Economics	Ciminelli & Garcia-Mandicó	When and how do business shutdowns work? Evidence from Italy's first COVID-19 wave	Excess daily deaths per 100,000	Incidence	Business Shutdowns	Italy		• Different Timing of Interventions • Adding/Removing Variables • Alternative Estimation Models • Alternative Variable Definitions
2021	JAMA Network Open	Ginther & Zambrana	Association of Mask Mandates and COVID-19 Case Rates, Hospitalizations, and Deaths in Kansas	Daily number of cases and deaths per county	Incidence	Mask Mandate	USA		
2021	JAMA Network Open	O'Donoghue	Association of University Student Gatherings With Community COVID-19 Infections Before and After the NCAA March Madness Tournament	New daily COVID-19 infections per 100,000	Incidence	Sport Games	USA		
2021	JAMA Network Open	Patel, et al.	Association of Simulated COVID-19 Vaccination and Nonpharmaceutical Interventions With Infections, Hospitalizations, and Mortality	Cumulative incidences of infections, hospitalizations, and deaths	Incidence	Vaccine efficacy and coverage with and without interventions	Simulation		• Varying Simulated Transmission Rates • Varying Simulated Mask Efficacy
2021	JAMA Network Open	Toumi, et al.	Association of Limited In-Person Attendance in US National Football League and National Collegiate Athletic Association Games With County-Level COVID-19 Cases	New COVID-19 cases per 100,000	Incidence	Sport Games	USA		• Different Timing of Interventions
2021	Nature Medicine	Ertem, et al.	The impact of school opening model on SARS-CoV-2 community incidence and mortality	Change in county-level incidence per 100,000	Log incidence	School Reopening	USA	Multivariate Poisson regression with robust standard errors	• Alternative Inference Methods
2021	PNAS	Berry, et al.	Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic	New cases or new deaths per 1,000,000	Incidence	Shelter-in-place	USA		• Different Timing of Interventions • Lagged Outcome or Covariates
2021	PNAS	Chernozhukov, Kasahara, & Schrimpf	The association of opening K–12 schools with the spread of COVID-19 in the United States: County-level panel data analysis	Approximating the case growth rate with the log difference in weekly confirmed cases	Log growth	School Reopening	USA	Motivated with an SIRD model	• Different Timing of Interventions • Adding/Removing Variables • Lagged Outcome or Covariates
2021	PNAS	Fetzer & Graeber	Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment	COVID-19 incidence per 100,000	Incidence	Contact Tracing	England		• Lagged Outcome or Covariates • Spatial Considerations • Alternative Functional Forms • Placebo Tests
2021	PNAS	Singh, et al.	Impacts of introducing and lifting nonpharmaceutical interventions on COVID-19 daily growth rate and compliance in the United States	Number of confirmed cumulative cases per 100,000 minus that on the prior day divided by that on the prior day	Incidence	Multiple Interventions	USA		• Different Timing of Interventions • Lagged Outcome or Covariates
2022	Health Affairs	Fitzpatrick, et al.	School Reopening And COVID-19 In The Community: Evidence From A Natural Experiment In Ontario, Canada	Natural log of the cumulative number of people testing positive on a given day minus that on the day before	Log incidence	School Reopening	Canada		• Different Timing of Interventions • Outliers • Adding/Removing Variables • Lagged Outcome or Covariates • Spatial Considerations

2022	Health Affairs	Nwadiuko & Bustamante	Little To No Correlation Found Between Immigrant Entry And COVID-19 Infection Rates In The United States	Lagged natural logarithm of new COVID-19 cases per 100,000	Log incidence	Immigrant Entry	USA	Outcome can be interpreted as percent change in lagged incidence per each additional entrant (i.e., semi-elasticity). The authors chose a semi-elastic instead of elastic (log-log) model, as the size might be more relevant. instead of percent change	• Outliers • Adding/Removing Variables • Lagged Outcome or Covariates • Alternative Functional Forms
2022	Health Economics	Goetz	Does providing free internet access to low-income households affect COVID-19 spread?	Number of cases per 1,000	Incidence	Free in-home Internet Access	Canada		• Different Timing of Interventions • Adding/Removing Variables • Alternative Functional Forms • Alternative Estimation Models
2022	JAMA Health Forum	Lichand, et al.	Association of COVID-19 Incidence and Mortality Rates With School Reopening in Brazil During the COVID-19 Pandemic	Log-Transformed new COVID-19 cases and deaths per 10,000	Log incidence	School Reopening	Brazil		• Alternative Estimation Models • Alternative Variable Definitions
2022	Nature Communication	Paetzold, et al.	Impacts of rapid mass vaccination against SARS-CoV2 in an early variant of concern hotspot	Weekly cases per 100,000	Incidence	Mass Vaccination	Austria		• Alternative Estimation Models • Placebo Tests
2022	NEJM	Cowger, et al.	Lifting Universal Masking in Schools - Covid-19 Incidence among Students and Staff	Weekly Covid-19 cases	Incidence	Mask Mandate	USA		• Different Timing of Interventions • Adding/Removing Variables
2022	PNAS	Diederichs, Isphording, & Pestel	Schools under mandatory testing can mitigate the spread of SARS-CoV-2	New confirmed SARS-CoV-2 infections	Incidence	School Reopening	Germany		• Different Timing of Interventions • Adding/Removing Variables
2022	The Lancet	McNamara, et al.	Estimating the early impact of the US COVID-19 vaccination programme on COVID-19 cases, emergency department visits, hospital admissions, and deaths among adults aged 65 years and older: an ecological analysis of national surveillance data	Relative change in incidence	Log incidence	Vaccine	USA	A negative binomial model is assumed because it is appropriate for count data	• Different Timing of Interventions • Alternative Reference Groups
2023	Health Economics	Agrawal, et al.	The impact of COVID-19 shelter-in-place policy responses on excess mortality	Excess deaths per 100,000	Incidence	Shelter-in-place	USA		• Different Timing of Interventions • Adding/Removing Variables • Alternative Functional Forms • Alternative Estimation Models • Alternative Variable Definitions
2023	Health Economics	Amodio, et al.	School-age vaccination, school openings and COVID-19 diffusion	Natural logarithm of COVID-19 cases, measured at times t-3 and t-4	Log growth	School Reopening	USA	Following the work by Chernozhukov, et al. (2021)	• Adding/Removing Variables • Lagged Outcome or Covariates
2023	Health Economics	Godøy & Grøtting	Implementation and spillovers of local non-pharmaceutical interventions	Daily updated records on polymerase chain reaction tests and test results	Incidence	Multiple Interventions	Norway		• Adding/Removing Variables • Lagged Outcome or Covariates • Alternative Estimation Models

B. Proofs

Proposition 1 (Expected incidence)**.** *Assuming an SIR data-generating process (Eq. 1) with initial conditions* $\{S_{d,0}, I_{d,0}, R_{d,0}\}$*, expected incidence at time* $t+1$ *can be written as:*

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

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

Proof. Given $\{I_{d,0}, S_{d,0}, R_{d,0}\}$, constant generation interval $\frac{1}{\gamma}$, and effective contact rates $(\beta_{d,t})$, we have per Eq. 1:

$$S_{d,t} = S_{t-1} - I^*_{d,t} = S_{d,0} - \sum_{i=1}^{t-1} I^*_{d,t}$$
$$I_{d,t} = (1-\gamma)\, I_{d,t-1} + I^*_{d,t}$$

We denote incidence at time $t+1$:

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

where $\mathbb{E}[\epsilon_t] = 0$ because $I^*_{d,t} \sim Pois\left(\beta_{d,t-1} I_{d,t-1}\frac{S_{d,t-1}}{N}\right)$.

Taking the expectation of both sides:

$$\mathbb{E}[I^*_{d,t+1}] = \mathbb{E}\left[\frac{\beta_{d,t}}{N} S_{d,t} I_{d,t} + \epsilon_{t+1}\right]$$

Noting that ϵ_{t+1} has mean 0 $\longrightarrow$

$$= \frac{\beta_{d,t}}{N}\mathbb{E}\left[S_{d,t} I_{d,t}\right]$$

Substituting $S_{d,t}, I_{d,t}$ from above $\longrightarrow$

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

Expanding product $\longrightarrow$

$$= \frac{\beta_{d,t}}{N}\mathbb{E}\left[S_{d,t-1} I^*_{d,t} + (1-\gamma) S_{d,t-1} I_{d,t-1} - (1-\gamma) I_{d,t-1} I^*_{d,t} - I^{*2}_{d,t}\right]$$

Taking expectations and noting $\mathbb{E}\left(S_{d,t-1} I_{d,t-1}\right) = \dfrac{\mathbb{E}\left[I^*_{d,t}\right]}{\beta_{d,t-1}}$ $\longrightarrow$

$$= \frac{\beta_{d,t}}{N}\left(\mathbb{E}\left[S_{d,t-1} I^*_{d,t}\right] + \frac{(1-\gamma)}{\beta_{d,t-1}}\mathbb{E}\left[I^*_{d,t}\right] - (1-\gamma)\mathbb{E}\left[I_{d,t-1} I^*_{d,t}\right] - \mathbb{E}\left[I^{*2}_{d,t}\right]\right)$$

Noting that $S_{d,t} = S_{d,0} - \sum_{i=1}^{t-1} I^*_{d,t}$ $\longrightarrow$

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

Collecting terms $\longrightarrow$

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

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

□

Proposition 2 (Expected incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1), with initial conditions* $\{S_0, I_0, R_0\}$*, for* $t \geq 1$*,*

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

Proof. We prove Proposition 2 by induction. For explication, we write $I^*_{d,t} = \beta_{d,t-1}I_{d,t-1}\frac{S_{d,t-1}}{N} + \epsilon_t$, where $\mathbb{E}\left[\epsilon_t\right] = 0$ because $I^*_{d,t} \sim Pois\left(\beta_{d,t-1}I_{d,t-1}\frac{S_{d,t-1}}{N}\right)$. Population size N is assumed constant (Eq. 1) over time, with $N = S_{d,0} + I_{d,0} + R_{d,0}$ for all t.

Base case $(t = 1)$:
We first take the expectation of $I_{d,1}$:

$$\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I_{d,t+1}\right] = \mathbb{E}\left[\lim_{S_{d,0}\to\infty}(1-\gamma)I_0 + \beta_{d,0}I_{d,0}\frac{S_{d,0}}{N} + \epsilon_{d,1}\right]$$

Noting that $\lim_{S_{d,0}\to\infty}\frac{S_{d,0}}{N} = \lim_{S_{d,0}\to\infty}\frac{S_{d,0}}{S_{d,0}+I_{d,0}+R_{d,0}} = 1$ and $\mathbb{E}[\epsilon_{d,1}] = 0 \longrightarrow$

$$= (1-\gamma+\beta_{d,0})I_{d,0}$$

We then take the expectation of incidence at time $t+1 = 2$:

$$\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I^*_{d,t+1}\right] = \mathbb{E}\left[\lim_{S_{d,0}\to\infty} I^*_{d,2}\right]$$

Substituting in from Eq. 1 $\longrightarrow$

$$= \mathbb{E}\left[\lim_{S_{d,0}\to\infty}\beta_{d,1}I_{d,1}\frac{S_{d,1}}{N} + \epsilon_{d,2}\right]$$

Noting that $\mathbb{E}[\epsilon_{d,2}] = 0 \longrightarrow$

$$= \beta_{d,1}\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I_{d,1}\frac{S_{d,1}}{N}\right]$$

Substituting in from Eq. 1 $\longrightarrow$

$$= \beta_{d,1}\mathbb{E}\left[\lim_{S_{d,0}\to\infty}\left((1-\gamma)I_{d,0} + \beta_{d,0}I_{d,0}\frac{S_{d,0}}{N} + \epsilon_{d,1}\right)\left(\frac{S_{d,0} - \beta_{d,0}I_{d,0}\frac{S_{d,0}}{N} - \epsilon_{d,1}}{N}\right)\right]$$

Rearranging the limit of products as a product of limits because both are finite $\longrightarrow$

$$= \beta_{d,1}\mathbb{E}\left[\lim_{S_{d,0}\to\infty}\left((1-\gamma)I_{d,0} + \beta_{d,0}I_{d,0}\frac{S_{d,0}}{N} + \epsilon_{d,1}\right)\lim_{S_{d,0}\to\infty}\left(\frac{S_{d,0}}{N} - \frac{\beta_{d,0}I_{d,0}S_{d,0}}{N^2} - \frac{\epsilon_{d,1}}{N}\right)\right]$$

Noting that $\lim_{S_{d,0}\to\infty}\frac{S_{d,0}}{N} = \lim_{S_{d,0}\to\infty}\frac{S_{d,0}}{S_{d,0}+I_{d,0}+R_{d,0}} = 1$ and $\mathbb{E}[\epsilon_{d,1}] = 0 \longrightarrow$

$$= \beta_{d,1}\left((1-\gamma)I_{d,0} + \beta_{d,0}I_{d,0}\right)$$

$$= \beta_{d,1}\left(1-\gamma+\beta_{d,0}\right)I_{d,0} = \beta_{d,1}\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I_{d,1}\right]$$

Induction step: Assume that:

$$\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I^*_{d,t}\right] = \beta_{d,t-1}\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I_{d,t-1}\right] = \beta_{d,t-1}\prod_{k=0}^{t-2}\left(1-\gamma+\beta_{d,k}\right) I_{d,0}$$

We will show that this implies that the same holds for the next time step, i.e.,

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

We have:

$$\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I^*_{d,t+1}\right] = \mathbb{E}\left[\lim_{S_{d,0}\to\infty} \beta_{d,t} I_{d,t}\frac{S_{d,t}}{N} + \epsilon_{d,t+1}\right]$$

Substituting in $\mathbb{E}[\epsilon_{d,t+1}] = 0 \longrightarrow$

$$= \beta_{d,t}\mathbb{E}\left[\lim_{S_{d,0}\to\infty} I_{d,t}\frac{S_{d,t}}{N}\right]$$

Substituting in from Eq. 1 $\longrightarrow$

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

Rearranging the limit of products as a product of limits because both are finite $\longrightarrow$

$$= \beta_{d,t}\mathbb{E}\left[\lim_{S_{d,0}\to\infty} \left((1-\gamma)I_{d,t-1} + I^*_{d,t}\right) \lim_{S_{d,0}\to\infty}\left(\frac{S_{d,0}}{N} - \frac{\sum_{j=1}^{t-1}\beta_{d,j}I_{d,j}S_{d,j}}{N^2} - \frac{\epsilon_{d,j}}{N}\right)\right]$$

Noting $\lim_{S_{d,0}\to\infty} \frac{S_{d,0}}{N} = 1$, $\mathbb{E}[\epsilon_{d,j}] = 0\,\forall j \longrightarrow$

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

Applying (1) and $\lim_{S_{d,0}\to\infty} \mathbb{E}\left[I^*_{d,t}\right] = \lim_{S_{d,0}\to\infty} \beta_{d,t-1}\mathbb{E}\left[I_{d,t-1}\right] \longrightarrow$

$$= \lim_{S_{d,0}\to\infty} \beta_{d,t}\left(1-\gamma+\beta_{d,t-1}\right)\mathbb{E}\left[I_{d,t-1}\right]$$

Applying that $\lim_{S_{d,0}\to\infty} \beta_{t-1}\mathbb{E}\left[I_{t-1}\right] = \beta_{d,t-1}\prod_{k=0}^{t-2}\left(1-\gamma+\beta_{d,k}\right)I_{d,0} \longrightarrow$

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

This completes the induction step.

Last, note that the dominated convergence theorem allows us to exchange limit and expectation because for all $S_{d,0}$, $I^*_{d,t}(S_{d,0})$ is stochastically dominated by $I^*_{d,t}{}' \sim Pois\left(\beta_{d,t-1}\prod_{k=0}^{t-2}\left(1-\gamma+\beta_{d,k}\right)I_{d,0}\right)$ (i.e., $I^*_{d,t}{}'$ from in a modified SIR process in which $I^*_{d,t}{}' \sim Pois(\beta_{d,t-1}I_{t-1})$) over all d, t.

□

Proposition 3 (Parallel trends: Incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) and an incidence model specification* $\left(Y_{d,t} = I^*_{d,t},\ g(y) = y\right)$*, the "infinite susceptible population" parallel trends assumption (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

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

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

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

$$\text{where } \mathbb{E}\left[I^*_{0,t_1}\right] = I_{d,0}\beta_{d,t_1-1}\prod_{k=0}^{t_1-2}\left(1-\gamma+\beta_{d,k}\right),$$

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

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 $\longrightarrow$

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

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

Substituting the above expression into the parallel trends condition, we obtain ‖:

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

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

□

Proposition 4 (Parallel trends: Log incidence (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) 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 (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

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

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

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

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 1 $\longrightarrow$

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

Dividing out common terms $\longrightarrow$

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

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

Therefore, the "infinite susceptible population" parallel trends assumption (Eq. 3) holds if and only if

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

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

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

□

‖ Note that in the special case of constant exponential growth, $Y_{d,t} = I_{d,0}\beta^t$, this condition reduces to $Y_{1,t_1}\left(\beta^{t_2-t_1} - 1\right) = Y_{0,t_1}\left(\beta^{t_2-t_1} - 1\right)$.

Proposition 5 (Parallel trends: Log growth (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1) 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 (Eq. 3) holds between* t_1 *and* t_2 *under the following conditions:*

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

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

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 1 $\longrightarrow$

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

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 defined in Eq. 3 completes the proof. □

Proposition 6 (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:*

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

.

Proof. Eq. 1 defines the average number of secondary infections per infected individual at time t as $\beta_{d,t}\frac{S_{d,t}}{N}$. Of individuals who become infected at time t, Eq. 1 also defines the fraction removed at each time step as $1-\gamma$. This gives us the effective reproduction number corresponding to the cohort becoming infectious at time t:

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

□

Proposition 7 (Parallel trends: Log R_t (infinite susceptible population))**.** *Assuming an SIR data-generating process (Eq. 1), 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 (Eq. 3) holds for all* t_1, t_2 *if and only if*

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

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

Proof. In Proposition 6, we defined $R_{d,t} = \sum_{j=t}^{\infty} (1-\gamma)^{j-t}\, \beta_{d,j} \frac{S_{d,j}}{N}$. Substituting this formulation of $R_{d,t}$ into the parallel trends assumption in Eq. 3, we obtain:

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

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

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

Substituting from Proposition 6 $\longrightarrow$

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

Taking limits $\longrightarrow$

$$\log\left(\frac{\sum_{j=t2}^{\infty}(1-\gamma)^{j-t_2}\,\beta_{1,j}}{\sum_{j=t_1}^{\infty}(1-\gamma)^{j-t_1}\,\beta_{1,j}}\right) = \log\left(\frac{\sum_{j=t_2}^{\infty}(1-\gamma)^{j-t_2}\,\beta_{0,j}}{\sum_{j=t_1}^{\infty}(1-\gamma)^{j-t_1}\,\beta_{0,j}}\right)$$

□

Proposition 8 (Parallel trends: Log β_t)**.** *Assuming an SIR data-generating process and a log-transformed effective reproduction number specification* $\left(Y_{d,t} = \log\left(\beta_{d,t}\right), g(\cdot) = \log(\cdot)\right)$*, the parallel trends assumption holds for all* $t_1, t_2 > t-1$ *if and only if*

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

Proof. Proposition 8 follows from a direct substitution of the outcome $Y_{d,t} = \beta_{d,t}$ into the log-transformed parallel trends assumption as defined in Eq. 3:

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

□

Proposition 9 (Estimation of $\beta_{d,t}$)**.** *Assuming an SIR data-generating process (Eq. 1), 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 likelihood (L) and log-likelihood (ℓ) functions can be defined:

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

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:

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

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

□

Supplement Proposition 1 (Log transformation). *Assume that Eq. 3 holds with a log link, such that:*

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

Then,

$$\mathbb{E}\Big[\log\left(Y_{1,t_2}(0)\right) - \log\left(Y_{1,t_1}(0)\right)\Big] - \mathbb{E}\Big[\log\left(Y_{0,t_2}(0)\right) - \log\left(Y_{0,t_1}(0)\right)\Big] \approx \left(\frac{Var(Y_{1,t_1})}{2\mathbb{E}[Y_{1,t_1}]^2} - \frac{Var(Y_{1,t_2})}{2\mathbb{E}[Y_{1,t_2}]^2}\right) - \left(\frac{Var(Y_{0,t_1})}{2\mathbb{E}[Y_{0,t_1}]^2} - \frac{Var(Y_{0,t_2})}{2\mathbb{E}[Y_{0,t_2}]^2}\right)$$

Proof. We can approximate $\mathbb{E}\left[\log(Y_{d,t})\right]$ with a second-order Taylor series expansion (38):

$$\mathbb{E}\left[\log(Y_{d,t})\right] \approx \log\left(\mathbb{E}\left[Y_{d,t}\right]\right) - \frac{Var(Y_{d,t})}{2\mathbb{E}\left[Y_{d,t}\right]^2}$$

Substituting this approximation for each term in $\mathbb{E}\Big[\log\left(Y_{1,t_2}(0)\right) - \log\left(Y_{1,t_1}(0)\right)\Big] - \mathbb{E}\Big[\log\left(Y_{0,t_2}(0)\right) - \log\left(Y_{0,t_1}(0)\right)\Big]$ completes the proof.

□

Corollary 1. *If the outcome is log incidence in Supplement Proposition 1, $Y_{d,t} = I^*_{d,t}$ with a log link, then,*

$$\mathbb{E}\Big[\log\left(Y_{1,t_2}(0)\right) - \log\left(Y_{1,t_1}(0)\right)\Big] - \mathbb{E}\Big[\log\left(Y_{0,t_2}(0)\right) - \log\left(Y_{0,t_1}(0)\right)\Big] \approx \left(\frac{1}{2\mathbb{E}[I^*_{1,t_1}]} - \frac{1}{2\mathbb{E}[I^*_{1,t_2}]}\right) - \left(\frac{1}{2\mathbb{E}[I^*_{0,t_1}]} - \frac{1}{2\mathbb{E}[I^*_{0,t_2}]}\right)$$

Proof. Per Eq. 1, we assume the incidence $I^*_{d,t}$ follows a Poisson distribution. As a result, the second-order term becomes:

$$\frac{Var\left(I^*_{d,t}\right)}{2\mathbb{E}\left[I^*_{d,t}\right]^2} = \frac{\mathbb{E}\left[I^*_{d,t}\right]}{2\mathbb{E}\left[I^*_{d,t}\right]^2} = \frac{1}{2\mathbb{E}\left[I^*_{d,t}\right]}$$

Substituting this into the result from Proposition 1 completes the proof. □

Supplement Proposition 2 (DiD with time-step aggregation). *Assume that the parallel trends assumption (Eq. 3) 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 pre-intervention time step t_1 and post-intervention time-step t_2, 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 aggregated parallel trends assumption (Eq. 5) also holds:

$$\log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_2\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_1\right]\right) = \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_2\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_1\right]\right),$$

where $\mathcal{T}_1$ and $\mathcal{T}_2$ denote aggregations of pre- and post-intervention time periods.

Proof. By imposing parallel trends (Eq. 3) for any pair of individual pre- and post-intervention time steps, we have for any t_1 and t_2,

$$\begin{aligned}
\log\left(\mathbb{E}\left[Y_{1,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t_1}(0)\right]\right) &= \log\left(\mathbb{E}\left[Y_{0,t_2}(0)\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t_1}(0)\right]\right) \iff \\
\log\left(\frac{\mathbb{E}\left[Y_{1,t_2}(0)\right]}{\mathbb{E}\left[Y_{1,t_1}(0)\right]}\right) &= \log\left(\frac{\mathbb{E}\left[Y_{0,t_2}(0)\right]}{\mathbb{E}\left[Y_{0,t_1}(0)\right]}\right) \iff \\
\frac{\mathbb{E}\left[Y_{1,t_2}(0)\right]}{\mathbb{E}\left[Y_{1,t_1}(0)\right]} &= \frac{\mathbb{E}\left[Y_{0,t_2}(0)\right]}{\mathbb{E}\left[Y_{0,t_1}(0)\right]} \iff \\
\mathbb{E}\left[Y_{1,t_2}(0)\right]\mathbb{E}\left[Y_{0,t_1}(0)\right] &= \mathbb{E}\left[Y_{0,t_2}(0)\right]\mathbb{E}\left[Y_{1,t_1}(0)\right] \qquad [(*)]
\end{aligned}$$

For the parallel trends condition to hold on the aggregated time periods $\mathcal{T}_1$ and $\mathcal{T}_2$, we want to show

$$\log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_2\right]\right) - \log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t \in \mathcal{T}_1\right]\right) = \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_2\right]\right) - \log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t \in \mathcal{T}_1\right]\right)$$

Note that if aggregation occurs by summing, rather than averaging, both the LHS and RHS are multiplied by the number of time steps, which does not affect the result. For each side of the above equation, the difference in log expected outcomes can be written as a sum of time steps:

$$
\begin{aligned}
&\log\left(\mathbb{E}\left[Y_{d,t}(0)\middle|t\in\mathcal{T}_2\right]\right)-\log\left(\mathbb{E}\left[Y_{d,t}(0)\middle|t\in\mathcal{T}_1\right]\right)\\
&=\log\left(\mathbb{E}\left[\sum_{t\in\mathcal{T}_2}Y_{d,t}(0)\right]\right)-\log\left(\mathbb{E}\left[\sum_{t\in\mathcal{T}_1}Y_{d,t}(0)\right]\right)\\
&=\log\left(\frac{\mathbb{E}\left[\sum_{t\in\mathcal{T}_2}Y_{d,t}(0)\right]}{\mathbb{E}\left[\sum_{t\in\mathcal{T}_1}Y_{d,t}(0)\right]}\right)\\
&=\log\left(\frac{\sum_{t\in\mathcal{T}_2}\mathbb{E}\left[Y_{d,t}(0)\right]}{\sum_{t\in\mathcal{T}_1}\mathbb{E}\left[Y_{d,t}(0)\right]}\right)
\end{aligned}
$$

Therefore, Eq. 5 holds for $\mathcal{T}_1$ and $\mathcal{T}_2$) if and only if

$$
\begin{aligned}
&\frac{\sum_{t\in\mathcal{T}_2}\mathbb{E}\left[Y_{1,t}(0)\right]}{\sum_{t\in\mathcal{T}_1}\mathbb{E}\left[Y_{1,t}(0)\right]}=\frac{\sum_{t\in\mathcal{T}_2}\mathbb{E}\left[Y_{0,t}(0)\right]}{\sum_{t\in\mathcal{T}_1}\mathbb{E}\left[Y_{0,t}(0)\right]}\iff\\
&\left(\sum_{t\in\mathcal{T}_2}\mathbb{E}\left[Y_{1,t}(0)\right]\right)\left(\sum_{t\in\mathcal{T}_1}\mathbb{E}\left[Y_{0,t}(0)\right]\right)=\left(\sum_{t\in\mathcal{T}_2}\mathbb{E}\left[Y_{0,t}(0)\right]\right)\left(\sum_{t\in\mathcal{T}_1}\mathbb{E}\left[Y_{1,t}(0)\right]\right)\iff\\
&\text{Distributing the sum}\longrightarrow\\
&\sum_{t_2\in\mathcal{T}_2}\sum_{t_1\in\mathcal{T}_1}\mathbb{E}\left[Y_{1,t_2}(0)\right]\mathbb{E}\left[Y_{0,t_1}(0)\right]=\sum_{t_2\in\mathcal{T}_2}\sum_{t_1\in\mathcal{T}_1}\mathbb{E}\left[Y_{0,t_2}(0)\right]\mathbb{E}\left[Y_{1,t_1}(0)\right],
\end{aligned}
$$

where the last equality is true because Eq. 3 holds for every pair of pre-intervention t_1 and post-intervention t_2. □

Corollary 2. *Assume that the parallel trends assumption (Eq. 3) 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 as in Supplement Proposition 2. Then the extended parallel trends assumption (Eq 6) holds on the transformed average differences between pre- and post-treatment outcomes across treated and untreated groups:*

$$
\log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t\in\mathcal{T}_2\right]\right)-\log\left(\mathbb{E}\left[Y_{1,t}(0)\middle|t\in\mathcal{T}_1\right]\right)=\log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t\in\mathcal{T}_2\right]\right)-\log\left(\mathbb{E}\left[Y_{0,t}(0)\middle|t\in\mathcal{T}_1\right]\right),
$$

where $\mathcal{T}_1$ and $\mathcal{T}_2$ are set of pre- and post-treatment time periods, respectively.

Proof. The proof follows directly from Supplement Proposition 2, with only the definitions of $\mathcal{T}_1$ and $\mathcal{T}_2$ changed. □

C. Figures

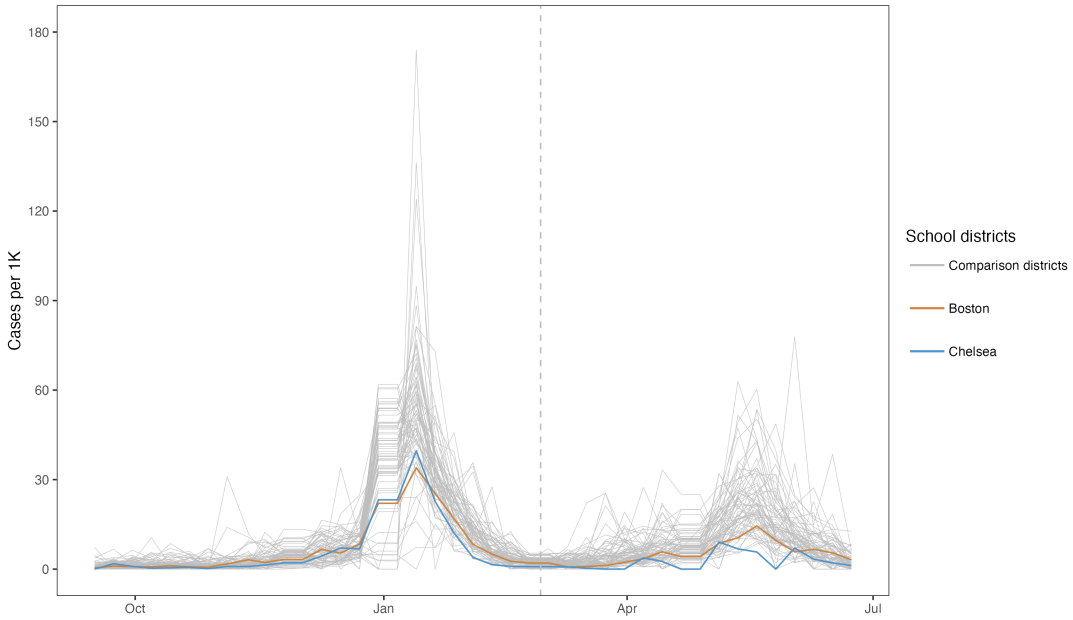

Fig. A1. COVID-19 cases per 1000 students and staff in 72 school districts in Massachusetts
Each line plots the case trajectory in one school district. The dashed vertical line represents the time at which the state-level mask mandate was lifted.

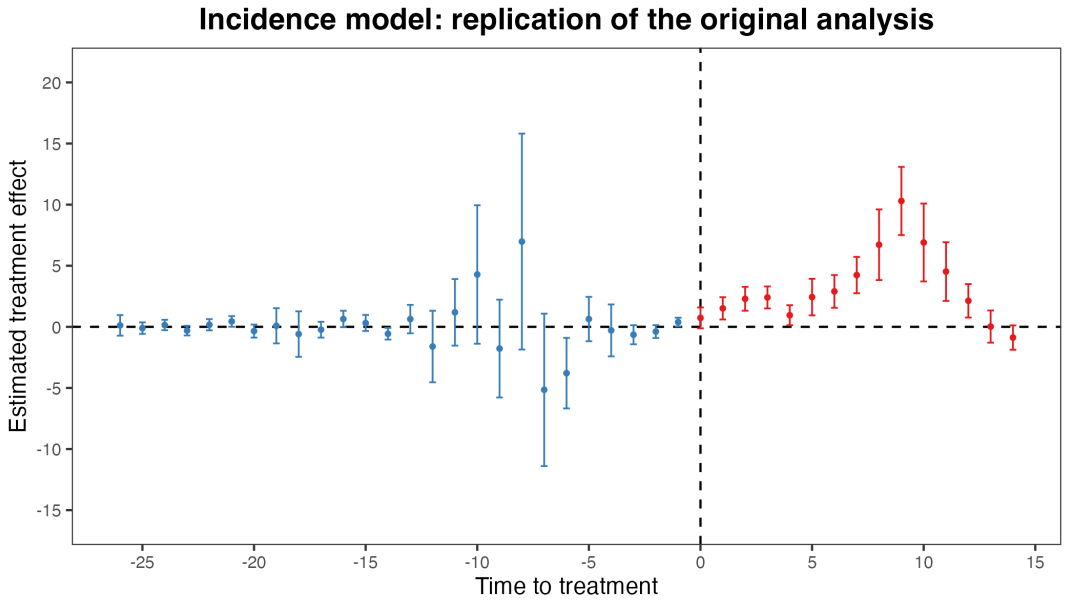

Fig. A2. Event study plot from replicating the original analysis Cowger et al. (5)
The estimated treatment effects with associated 95% confidence intervals obtained from fitting the incidence specification, which is the model specification in the original analysis, are plotted against the time to treatment.

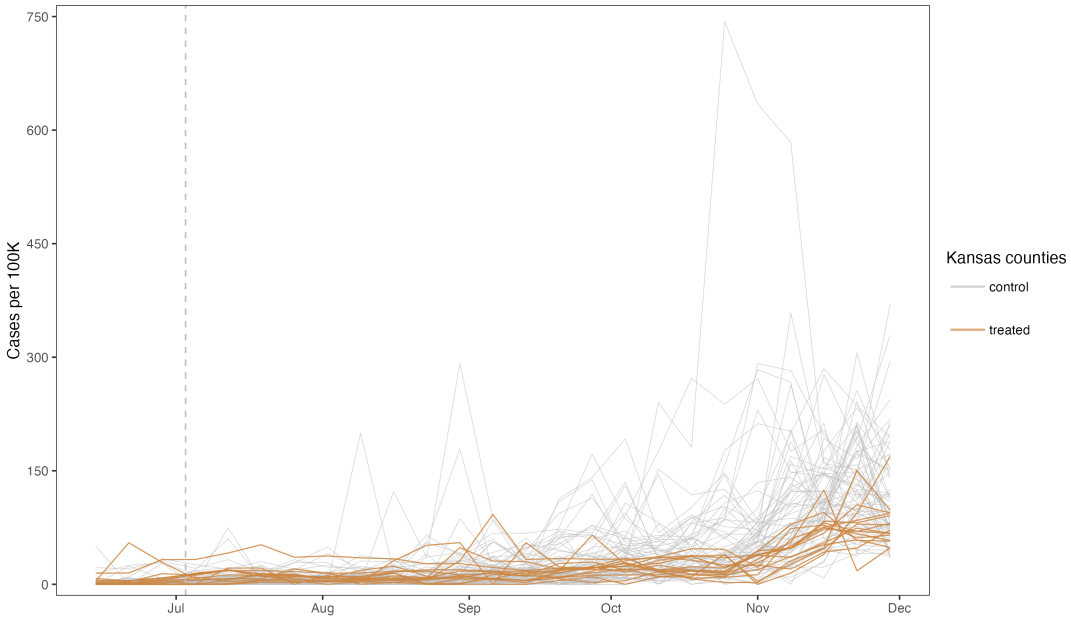

Fig. A3. COVID-19 cases per 100,00 population in Kansas

Each line plots the case trajectory in one county in Kansas. The dashed vertical line represents the time at which the executive order took effect.

D. Tables

Table A1. Simulation parameters

Parameter	Description	Values
N	Total number of units	50
N_1	Number of treated units	25
pop	Population size for each unit	One of $\{5,000,\ 10,000\}$
T	Total number of weeks	17
T_{burnin}	Number of weeks in the burn-in period	5
T_0	Number of weeks in the pre-intervention period	4
I_0	Number of initial infections at time 0	100
$\frac{1}{\gamma}$	Generation interval	10 days
β	Effective contact rate (constant over time)	One of $\{0.100,\ 0.115\}$
ϕ	Ratio in the effective contact rates between the treated and control groups	One of $\{1.0,\ 1.1\}$
δ	Effect size	One of $\{0.70, 0.80, 0.90, 0.95, 1.00, 1.05, 1.10, 1.20, 1.30\}$
α	Statistical significance level	0.05

Table A2. Results from re-analyzing the effect of removing school mask mandates on COVID-19 cases in Massachusetts using Callaway and Sant'Anna estimator

Outcome specification	Treatment effect[1]	Average marginal effect[2]
Incidence	47.4	47.4
Log incidence	2.0	10.0
Log growth	-2.2	-179.5

[1] Treatment effect refers to the point estimate in coefficients obtained from OLS (incidence model) or Poisson regression (log incidence and log growth models) [2] Average treatment effects are imputed as described in Section **Average marginal effects** and Appendix E.

E. Algorithms

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

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 1. 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: $ME_i = \sum_{t \in \mathcal{T}_1} \left(I^*_{d,t} - \widehat{I^*_{d,t}}(0) \right)$
4. The AME is the average of the calculated differences over all treated units: $AME = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1} (ME_i)$

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 (36). 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 as follows:

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 average difference in projected infections over the post-intervention period between fitted trajectories: $ME_i = \sum_{t \in \mathcal{T}_1} \left(I^*_{d,t} - \widehat{I^*_{d,t}}(0) \right)$, $AME = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1} (ME_i)$.

F. SEIR framework

We summarize the SEIR framework using the following equations:

$$\begin{aligned}
\textbf{Susceptible:}\ & S_{d,t+1} = S_{d,t} - \left(\beta_{d,t} I_{d,t} \frac{S_{d,t}}{N} + \epsilon_{d,t+1} \right) \\
\textbf{Exposed:}\ & E_{d,t+1} = (1-\psi) E_{d,t} + \left(\beta_{d,t} I_{d,t} \frac{S_{d,t}}{N} + \epsilon_{d,t+1} \right) \\
\textbf{Infectious:}\ & I_{d,t+1} = (1-\gamma) I_{d,t} + \psi E_{d,t} \\
\textbf{Recovered:}\ & R_{d,t+1} = R_{d,t} + \gamma I_{d,t}
\end{aligned}$$

Compared with an SIR model specified in Eq. 1, the only additional parameter introduced here is ψ, which in our analysis is assumed to be a constant rate of infectious given exposure.

G. 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, proposed by Cameron et al. (23). The more widely-used wild cluster bootstrap perturbs the residual distribution for each bootstrap replicate based on a cluster-level random variable with mean 0 and variance 1, usually drawn from a Rademacher distribution (23, 24). Although this technique performs well with a small number of clusters, it requires a symmetric distribution in the residuals with mean 0, which is not satisfied by a Poisson generalized linear model (GLM). More broadly, this idea of perturbing distributions in bootstrap samples can be applied to the score contributions in the context of GLMs (23). Given any maximum likelihood estimation process, the score contribution for cluster c can be computed 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. In each bootstrap replicate, we re-weight the score distribution based on an auxiliary cluster-level random variable with mean 0 and variance 1, and 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.