Social distancing is widely acknowledged as an effective public health policy combating the novel coronavirus. But extreme forms of social distancing like isolation and quarantine have costs and it is not clear how much social distancing is needed to achieve public health effects. In this article, we develop a design-based framework to test the causal null hypothesis and make inference about the dose-response relationship between reduction in social mobility and COVID-19 related public health outcomes. We first discuss how to embed observational data with a time-independent, continuous treatment dose into an approximate randomized experiment, and develop a randomization-based procedure that tests if a structured dose-response relationship fits the data. We then generalize the design and testing procedure to a longitudinal setting, and apply them to investigate the effect of social distancing during the first phased reopening in the United States on public health outcomes using data compiled from Unacast™, the United States Census Bureau, and the County Health Rankings and Roadmaps Program. We rejected a primary analysis null hypothesis that stated the social distancing from April 27, 2020, to June 28, 2020, had no effect on the COVID-19-related death toll from June 29, 2020, to August 2, 2020 ($p$-value < 0.001), and found that it took more reduction in mobility to prevent exponential growth in case numbers for non-rural counties compared to rural counties.

1. Introduction.

1.1. Social distancing, a pilot study, and dose-response relationship. Social distancing is widely acknowledged as one of the most effective public health strategies to reduce transmission of the novel coronavirus (Lewnard and Lo, 2020). There seemed to be ample evidence from China (Lau et al., 2020) and Italy (Sjödin et al., 2020) that a strict lockdown and practice of social distancing could have a substantial effect on reducing disease transmission, but social distancing has economic, psychological and societal costs (Acemoglu et al., 2020; Atalan, 2020; Grover et al., 2020; Sheridan et al., 2020; Venkatesh and Edirappuli, 2020). How much social distancing is needed to achieve the desired public health effect? In this article, we measure the level of social distancing using data on daily percentage change in total distance traveled compared to the pre-coronavirus level (data compiled and made available by Unacast™) and investigate the causal relationship between social distancing and COVID-related public health outcomes.

We conducted a pilot study in March to investigate the effect of social distancing during the first week of President Trump’s 15 Days to Slow the Spread campaign (March 16–22,
2020) on the influenza-like illness (ILI) percentage two and three weeks later. We tested the causal null hypothesis and found some weak evidence \( (p\text{-value} = 0.08) \) that better social distancing had an effect on ILI percentage three weeks later. In Supplementary Material A, we described in detail our pilot study. A protocol of the design and analysis was posted on arXiv (https://arxiv.org/abs/2004.02944) before outcome data were available and analyzed.

In addition to the causal null hypothesis, the “dose-response relationship” between the degree of social distancing and potential public health outcomes under various degrees of social distancing is also of great interest. Infectious disease experts expressed sentiments that the effect of social distancing on public health outcomes might be small or even negligible under a small degree of social distancing, but much more substantial under a large degree of social distancing. In an interview with the British Broadcasting Corporation (BBC Radio 4, 2020), director of the National Institute of Allergy and Infectious Diseases (NIAID), Dr. Anthony S. Fauci said:

“We never got things down to baseline where so many countries in Europe and the UK and other countries did – they closed down to the tune of about 97 percent lockdown. In the United States, even in the most strict lockdown, only about 50 percent of the country was locked down. That allowed the perpetuation of the outbreak that we never did get under very good control”.

Perhaps Dr. Fauci was proposing a hypothesis that the treatment dose, i.e., level of social distancing, played a very important role, and the causal effect of social distancing as a public health strategy combating coronavirus transmission is likely to be very different depending on the extent to which it is practiced (see, e.g., Gelfand et al. (2021)). We would like to formalize and test this dose-response relationship hypothesis.

1.2. Reopening, causal null hypothesis, and dose-response kink model. Starting late April and early May, many states in the U.S. started phased reopening. States and local governments differed in their reopening timelines; people in different states and counties also differed in their mobility during the process: some ventured out; some continued to stay at home as much as possible. The right panel of Figure 1 plots the 7-day rolling average of percentage change in total distance traveled of all counties in the U.S., from mid-March to late May. It is evident that as many counties started to ease social distancing measures, we saw less reduction in distance traveled; in fact, in many counties, distance traveled started to return to and even supersede the pre-coronavirus level. The phased reopening was followed by quickly deteriorating public health outcomes in July; see the left panel of Figure 1.

In this article, we leverage the county-level social mobility data since phased reopening to study the relationship between social mobility and its effect on public health outcomes. Let \( z_{t_0:T} \) be a longitudinal measurement of change in social mobility in county \( n \) from period \( t_0 \) to \( T \), and let \( Y_{n,T}(z_{t_0:T}) \) be county \( n \)'s potential public health outcome at time \( T \) under the social mobility trajectory \( z_{t_0:T} \), e.g., the number of patients succumbing to the COVID-19 at time \( T \). Our first scientific query is about the causal null hypothesis: Had the social mobility trajectory changed from \( z_{t_0:T} \) to \( z'_{t_0:T} \), would the potential public health outcome at time \( T \) change at all? In other words, does \( Y_{n,T}(z_{t_0:T}) = Y_{n,T}(z'_{t_0:T}) \) hold for all \( z_{t_0:T} \neq z'_{t_0:T} \)? Suppose that we have enough evidence from observational data to reject this causal null hypothesis, our second query then is about the dose-response relationship between the level of social distancing and its effect on the potential public health outcome. For instance, let \( z \) capture some aggregate dose of \( z_{t_0:T} \), e.g., the average reduction in social mobility from \( t_0 \) to \( T \), and \( z^* \) a reference dose level. One dose-response relationship (among many other candidates) is the following dose-response kink model (see Figure 2):

\[
H^K_0 : \quad Y_{n,T}(z) = Y_{n,T}(z^*), \quad z \leq \tau,
\]
\[
Y_{n,T}(z) - Y_{n,T}(\tau) = \beta(z - \tau), \quad z > \tau,
\]

(1)
Fig 1: Left panel: 7-day average COVID-19 death toll in the United States in 2020. The phased reopening since early May seemed to have contributed to a major hike in the death toll in July. Right panel: county-averaged 7-day rolling average (black solid line), middle 50% (dark shade), and middle 90% (light shade) of percentage change in total distance traveled, e.g., −0.35 corresponds to 35% reduction in total distance traveled compared to the pre-coronavirus period. The first week of 15 Days to Slow the Spread campaign (March 16-22) is marked in red and the first week of reopening in blue.

for all \( n \) and some \( \tau \) and \( \beta \). Model (1) states that the potential public health outcome (e.g., daily death toll, test positivity rate, etc) at time \( T \) would remain unchanged as the potential outcome under the reference level when the aggregate dose \( z \) is less than a certain threshold \( \tau \), and then increases at a rate proportional to how much \( z \) exceeds the threshold. Model (1) succinctly captures two key features policy makers may be most interested in: \( \tau \) the minimum dose that “activates” the treatment effect, and \( \beta \) how fast the potential outcome changes as the dose changes.

Our analysis in this article complements standard analyses based on epidemiological models, e.g., the SIR (susceptible-infected-recovered) compartment models (Brauer and Castillo-Chavez, 2012). The standard approach focuses on how the public health outcome trajectory
evolves over time. Our approach doesn’t provide information on this, but provides information on the causal effect without needing strong assumptions on how the public health outcome evolves over time. A parsimonious dose-response relationship does not preclude nonlinear infectious disease dynamics, e.g., those based on the SIR and other models; moreover, our primary inferential target, the causal null hypothesis, does not impose any restriction on the infectious disease mechanism.

1.3. Our contribution. We have three goals in this article. First, we propose a simple, model-free randomization-based procedure that tests if a causal null hypothesis or a structured dose-response relationship, e.g., the dose-response kink model, fits the data in a static setting with a time-independent, continuous treatment dose. To be specific, an empirical researcher posits a structured dose-response relationship that she finds scientifically meaningful, parsimonious, and flexible enough to describe data at hand; our developed procedure can then be applied to test if such a postulated dose-response relationship is sufficient to describe the causal relationship. If the hypothesis is rejected, empirical researchers are then advised to re-examine the scientific theory underpinning the postulated model; otherwise, the model seems a good starting point for data analysis. In this way, our method can be deemed as a model-free “diagnostic test” for a dose-response relationship, and more broadly a test of the underlying scientific theory.

In our application, the treatment and outcome are both longitudinal. Our second goal is to generalize the proposed design and testing procedure to the longitudinal setting under appropriate assumptions. We define a notion of cumulative dose for a time-varying treatment dose trajectory, and discuss how to embed observational longitudinal data into an approximate randomized controlled trial in order to permute two treatment trajectories. Finally, we closely examine our assumptions in the context of an infectious disease transmission mechanism and apply the developed design and testing procedure to characterize the dose-response relationship between reduction in social mobility and public health outcomes during the reopening phases in the U.S. using county-level data we compiled from sources including Unacast™, the United States Census Bureau, and the County Health Rankings and Roadmaps Program (Remington, Catlin and Gennuso, 2015).

The rest of the article is organized as follows. Section 2 and 3 study how to investigate a dose-response relationship using nonbipartite matching in a static setting. Section 4 incorporates interference and considers the spillover effects. Section 5 extends the method to longitudinal studies and describes the design of the case study. Section 6 presents results and extensive sensitivity analyses. Section 7 concludes with a discussion.

2. Investigating the dose-response relationship via nonbipartite matching.

2.1. Observational data with a continuous treatment dose in a static setting. Suppose there are \( N = 2I \) units, indexed by \( n = 1, \ldots, N \). Each unit is associated with a vector of observed covariates \( X_n \), an observed treatment dose assignment \( Z_{n}^{\text{obs}} \), and an observed outcome \( Y_{n}^{\text{obs}} \). The observed covariates \( X_n \) are collected before the treatment assignment and not affected by the treatment. Let \( Z_n \) be the treatment dose assignment of unit \( n \), and \( Z \) the set of all possible treatment doses. The cardinality of \( Z \) is \( |Z| = 2 \) for a binary treatment and an infinite number for a continuous treatment.

Let \( Y_n(z) \) be the potential outcome that unit \( n \) exhibits under the dose assignment \( z \) assuming no interference among units (Rubin, 1980, 1986). Each unit \( n \) is associated with a possibly infinite array of potential outcomes \( \{Y_n(z)\}_z \). We assume consistency so that \( Y_{n}^{\text{obs}} = Y_n(Z_{n}^{\text{obs}}) \). A causal estimand is necessarily a contrast between potential outcomes. Each unit \( n \) is associated with a collection of unit-level causal effects \( \{Y_n(z) - Y_n(z')\}_{z,z'} \).
Table 1 is called a \textit{science table} in the literature (Rubin, 2005), and summarizes information regarding these \(N\) units, where we let \(Z = \mathbb{N}\) for ease of exposition. In a causal inference problem, the fundamental estimands of interest are the arrays of potential outcomes in Table 1; the task of uncovering the arrays of potential outcomes is challenging because only one of the potentially infinite array of potential outcomes for each unit is actually observed.

|   | Covariates \(X\) | Observed \(Z\) | Potential Outcomes | Unit-level Causal Effects | Summary Causal Effects |
|---|---|---|---|---|---|
| 1 | \(X_1\) | \(Z_{1x}^{0w}\) | \(Y_1(1)\) \(\ldots\) \(Y_1(z')\) \(\ldots\) | \((Y_1(z) - Y_1(z'))_{z',z}\) | Summarize dose-response relationship for a common set of units |
| \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) |
| \(N\) | \(X_N\) | \(Z_{Nx}^{0w}\) | \(Y_N(1)\) \(\ldots\) \(Y_N(z')\) \(\ldots\) | \((Y_N(z) - Y_N(z'))_{z',z}\) | \(f_n(z; z^*, \theta_n)\) |

One unique feature of a continuous treatment dose is that the unit-level causal effect is an infinite set of comparisons between any two potential outcomes \(Y_n(z)\) and \(Y_n(z')\), unlike with a binary treatment where the unit-level causal effect unambiguously refers to a comparison between \(Y_n(1)\) and \(Y_n(0)\). Let \(z^*\) be an arbitrary reference dose. Observe that \(Y_n(z) - Y_n(z^*) = Y_n(z) - Y_n(z^*) - \{Y_n(z') - Y_n(z^*)\}\), and the collection of contrasts \(\{Y_n(z) - Y_n(z^*)\}_{z}\) is sufficient in summarizing all pairwise comparisons of potential outcomes. With a binary treatment, a “summary causal effect” (Rubin, 2005) is defined as a comparison between \(Y_n(1)\) and \(Y_n(0)\) over the same collection of units. With a continuous treatment dose, we first summarize the causal effects with a “unit-level dose-response relationship” for each unit \(n\). For example, one simple relationship states that \(Y_n(z) - Y_n(z^*) = \beta(z - z^*)\) for all \(z\); in words, for unit \(n\), the causal effect when comparing dose \(z\) to the reference dose \(z^*\) is proportional to the difference between \(z\) and \(z^*\). We may then summarize such unit-level dose-response relationships for a collection of units. For example, one such summary may state that a structured dose-response relationship \(f(z; z^*, \theta)\) holds for all counties in the U.S., and is represented by the following null hypothesis:

\[
H_0^1: Y_n(z) - Y_n(z^*) = f(z; z^*, \theta),
\]

for all counties indexed by \(n\) and some \(\theta\). We first develop a simple, randomization-based testing procedure to assess hypotheses of the form \(H_0^1\). In randomization inference, the potential outcomes (i.e., the infinite collection \(\{Y_n(z)\}_{z,n}\) in Table 1) are held fixed and the only probability distribution entering statistical inference is the randomization distribution that describes the treatment dose assignment; therefore, the key step here is to properly embed the observational data into an approximately randomized experiment via statistical matching (Rosenbaum, 2002, 2010). The work most relevant to our development is Ding, Feller and Miratrix (2016), who studied testing the existence of treatment effect variation in a randomized controlled trial with a time-independent binary treatment.

2.2. \textit{Embedding observational data with a time-independent, continuous treatment into an as-if randomized experiment via nonbipartite matching.} In a randomized controlled experiment, physical randomization creates “the reasoned basis” for drawing causal inference (Fisher, 1935). In the absence of physical randomization as with retrospective observational data, one strategy is to use statistical matching to embed observational data into a hypothetical randomized controlled trial (Rosenbaum, 2002, 2010; Rubin, 2007; Ho et al., 2007; Stuart, 2010; Bind and Rubin, 2019) by matching subjects with the same (or at least very
similar) estimated propensity score or observed covariates and designing two groups that are well-balanced in observed covariates.

One straightforward design to handle observational data with a continuous treatment is to dichotomize the continuous treatment based on some prespecified threshold, create a binary treatment out of the dichotomization scheme, and match on the binary treatment. This strategy is often seen in empirical research, probably because of its simplicity; however, dichotomizing the continuous treatment inevitably censors the rich information contained in the original dose and prevents researchers from studying the dose-response relationship.

To address this limitation, Lu et al. (2001, 2011) proposed optimal nonbipartite matching. In a nonbipartite matching, units with similar observed covariates but different treatment doses are paired. Suppose there are \( N = 2I \) units, e.g., counties in the U.S. in our application. In the design stage, distances \( \{\delta_{ij}, \ i, j = 1, \ldots, N\} \) are calculated between any two units and a \( N \times N \) distance matrix is constructed (Lu et al., 2001, 2011; Baiocchi et al., 2010). Some commonly used distances \( \delta_{ij} \) include the Mahalanobis distance between observed covariates \( X_i \) and \( X_j \) and the rank-based robust Mahalanobis distance. Researchers may further modify the distance to incorporate specific design aspects of the study. For instance, in a study involving effect modification, researchers are advised to match exactly or near-exactly on the effect modifier (Rosenbaum, 2005), e.g., the geographic location of the county, and this can be pursued by adding a large penalty to \( \delta_{ij} \) if county \( i \) and \( j \) are not from the same geographic region.

An optimal nonbipartite matching algorithm then divides these \( N = 2I \) units into \( I \) non-overlapping pairs that minimize the total within-matched-pair distance.

Suppose that we have formed \( I \) matched pairs of \( 2 \) units so that index \( ij \) uniquely identifies a unit, \( i = 1, \ldots, I, \ j = 1, 2 \). Let \( M_i^{\text{obs}} = \max(Z_{i1}^{\text{obs}}, Z_{i2}^{\text{obs}}) \) and \( m_i^{\text{obs}} = \min(Z_{i1}^{\text{obs}}, Z_{i2}^{\text{obs}}) \). We define the following two potential outcomes for each unit \( ij \):

\[
\begin{align*}
Y_{Tij} &\overset{\Delta}{=} Y_{ij}(M_i^{\text{obs}}), \\
Y_{Cij} &\overset{\Delta}{=} Y_{ij}(m_i^{\text{obs}}),
\end{align*}
\]

where we abuse the notation and use subscripts \( T \) and \( C \) to denote the potential outcomes under the maximum and minimum of two observed doses within each matched pair, respectively (Rosenbaum, 1989; Heng et al., 2019).

Write \( \mathcal{F} = \{X_{ij}, Y_{Tij}, Y_{Cij}\}_{i,j} \), where \( Y_{Tij} \) and \( Y_{Cij} \) are defined in (2), \( Z_{\text{max}}^{\text{obs}} = (M_1^{\text{obs}}, \ldots, M_I^{\text{obs}}) \), and \( Z_{\text{min}}^{\text{obs}} = (m_1^{\text{obs}}, \ldots, m_I^{\text{obs}}) \). As always in randomization inference (Rosenbaum, 2002, 2010; Ding, Feller and Miratrix, 2016), we condition on observed covariates, potential outcomes, and observed dose assignments, i.e., we do not model \( X \) or the potential outcomes, and rely on the treatment assignment mechanism to draw causal conclusions. The law describing the treatment dose assignment in matched pair \( i \) is

\[
\pi_{i1} = P(Z_{i1} = M_i^{\text{obs}}, Z_{i2} = m_i^{\text{obs}} \mid \mathcal{F}, Z_{\text{max}}^{\text{obs}}, Z_{\text{min}}^{\text{obs}}),
\]

and \( \pi_{i2} = 1 - \pi_{i1} \). In an ideal randomized experiment, experimenters use physical randomization (e.g., coin flips) to ensure \( \pi_{i1} = \pi_{i2} = 1/2 \): for matched pair \( i \) with two treatment doses \( Z_{i1}^{\text{obs}} \) and \( Z_{i2}^{\text{obs}} \), a fair coin is flipped; if the coin lands heads, the first unit is assigned \( Z_{i1}^{\text{obs}} \) and the second unit \( Z_{i2}^{\text{obs}} \), and vice versa if the coin lands tails. The design stage of an observational study aims to approximate this ideal (yet unattainable) hypothetical experiment by matching units with similar covariates \( X \) so that \( \pi_{i1} \approx \pi_{i2} \) after matching. In this way, nonbipartite matching embeds observational data with a continuous treatment dose into a randomized experiment; this induced randomization scheme will serve as our “reasoned basis” for inferring any causal effect including a dose-response relationship. As is always true with retrospective observational studies, a careful design may alleviate, but most likely never eliminate bias due to the residual imbalance in \( X \) or unmeasured confounding variables.

The departure from randomization, i.e., \( \pi_{i1} \neq \pi_{i2} \), is investigated via a sensitivity analysis (Rosenbaum, 1989, 2002, 2010).
3. Randomization inference for a dose-response relationship.

3.1. Randomization inference for \( \tau = \tau_0 \) and \( \beta = \beta_0 \) in the dose-response kink model.

We consider testing the dose-response kink model for a fixed \( \tau = \tau_0 \) and \( \beta = \beta_0 \) for all units:

\[
H_{0, \text{kink}}^{\tau_0, \beta_0} : Y_{ij}(z) = Y_{ij}(z^*) , \quad z \leq \tau_0 ,
\]
\[
Y_{ij}(z) - Y_{ij}(\tau_0) = \beta_0 (z - \tau_0) , \quad z > \tau_0 ,
\]

where \( z^* \) is some reference dose level, e.g., the minimum dose level. Under \( H_{0, \text{kink}}^{\tau_0, \beta_0} \), the entire dose-response relationship for subject \( ij \) is known up to \( Y_{ij}(z^*) \). Fortunately, we do observe one point on the dose-response curve, namely \( Y_{ij}(Z_{ij}^{\text{obs}}) \); hence, the entire dose-response curve is fixed, and both potential outcomes \( Y_{ij}(M_{ij}^{\text{obs}}) \) and \( Y_{ij}(M_{ij}^{\text{obs}}) \) can be imputed for each unit. In matched pair \( i \), for the unit with \( Z_{ij}^{\text{obs}} = m_{ij}^{\text{obs}} \), its potential outcome under \( m_{ij}^{\text{obs}} \) is the observed outcome \( Y_{ij}^{\text{obs}} \) and under \( M_{ij}^{\text{obs}} \) is

\[
Y_{ij}(M_{ij}^{\text{obs}}) = \begin{cases} 
Y_{ij}^{\text{obs}} & M_{ij}^{\text{obs}} \leq \tau_0 , \\
Y_{ij}^{\text{obs}} + \beta_0 \times (M_{ij}^{\text{obs}} - \tau_0) , & M_{ij}^{\text{obs}} \leq \tau_0 \text{ and } M_{ij}^{\text{obs}} > \tau_0 ; \\
Y_{ij}^{\text{obs}} & M_{ij}^{\text{obs}} > \tau_0 ,
\end{cases}
\]

For the unit with \( Z_{ij}^{\text{obs}} = M_{ij}^{\text{obs}} \), its potential outcome under \( M_{ij}^{\text{obs}} \) is the observed outcome \( Y_{ij}^{\text{obs}} \) and under \( m_{ij}^{\text{obs}} \) is

\[
Y_{ij}(m_{ij}^{\text{obs}}) = \begin{cases} 
Y_{ij}^{\text{obs}} & m_{ij}^{\text{obs}} \leq \tau_0 , \\
Y_{ij}^{\text{obs}} - \beta_0 \times (m_{ij}^{\text{obs}} - \tau_0) , & m_{ij}^{\text{obs}} \leq \tau_0 \text{ and } m_{ij}^{\text{obs}} > \tau_0 ; \\
Y_{ij}^{\text{obs}} & m_{ij}^{\text{obs}} > \tau_0 ,
\end{cases}
\]

Table 2 illustrates the imputation scheme by imputing the missing potential outcome (represented by “?’” in the table) under the null hypothesis \( H_{0, \text{kink}}^{\tau_0, \beta_0} \) with \( \tau_0 = 0.3 \) and \( \beta_0 = 1 \).

**Table 2**

<imsetable>

| Units | Observed Dose \( Z_{ij}^{\text{obs}} \) | Observe One Potential Outcome | Imputed Potential Outcomes |
|-------|----------------------------------|-------------------------------|-----------------------------|
|       | \( Y_{ij}(m_{ij}^{\text{obs}}) \) | \( Y_{ij}(M_{ij}^{\text{obs}}) \) | \( Y_{ij}(m_{ij}^{\text{obs}}) \) | \( Y_{ij}(M_{ij}^{\text{obs}}) \) |
| 11    | 0.2                             | ?                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
| 12    | 0.4                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
| 21    | 0.9                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
| 22    | 2.2                             | ?                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
| 31    | 1.4                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
| 32    | 1.9                             | ?                             | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) |
|       |                                 |                               |                             |                             |
| 11    | \( Z_{ij} \)                     | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) | Impute according to scheme (3) |
| 12    | \( Z_{ij} \)                     | \( Y_{ij}^{\text{obs}} \) | \( Y_{ij}^{\text{obs}} \) | Impute according to scheme (4) |

Let \( ij' \) be the unit with dose \( m_{ij'}^{\text{obs}} \) in matched pair \( i \), \( Y_{ij'}^{\text{obs}} = \{ Y_{ij'}^{\text{obs}} \} \), the observed outcomes of units with dose \( m_{ij'}^{\text{obs}} \) in each pair, and \( \hat{F}_{\text{min}} \) its empirical distribution function.
Analogously, let $ij''$ be the unit with dose $M_i^{obs}$ and $Y_{max}^{obs} = \{Y_{ij''}^{obs}\}_i$ the observed outcomes of units with dose $M_i^{obs}$ in each pair. For each $Y_{ij''}^{obs} \in Y_{max}^{obs}$, define its transformed outcome $\tilde{Y}_{ij''}^{obs} = Y_{ij''}^{obs}(m_i^{obs})$ according to (4). Let $\tilde{Y}_{max}^{obs} = \{\tilde{Y}_{ij''}^{obs}\}_i$ be the collection of transformed outcomes, and $\hat{F}_{max}^{tr}$ its empirical distribution function. The null hypothesis $H_{0, kink}^{\tau_0, \beta_0}$ can then be tested by comparing the following Kolmogorov–Smirnov-type (KS) test statistic

$$t_{KS}(\tau_0, \beta_0) = \sup_y |\hat{F}_{min}(y) - \hat{F}_{max}^{tr}(y)|$$

evaluated at the observed data to a reference distribution generated by imputing the science outcomes, and $\beta_0 = 0.5$. We test the null hypothesis $H_{0, kink}^{\tau_0, \beta_0}$ with $\tau_0 = 1$ and $\beta_0 = 0.5$ using the test statistic (5). The left panel of Figure 3 plots the empirical distribution $\hat{F}_{min}(y)$ (blue) and $\hat{F}_{max}^{tr}(y)$ (red), and $t_{KS}(1, 0.5) = 0.07$ for the observed data. Instead of enumerating all $2I = 2^{200}$ possible treatment dose assignments, we draw with replacement 100,000 samples from all $2^{200}$ possible configurations. The right panel of Figure 3 plots the reference distribution based on these 100,000 samples. In this way, a Monte Carlo $p$-value equal to 0.61 is obtained in this simulated dataset and the null hypothesis $H_{0, kink}^{\tau_0, \beta_0}$ with $\tau_0 = 1$ and $\beta_0 = 0.5$ is not rejected.

![Empirical distribution function](image1)

![Density](image2)

Fig 3: An illustrative example. $I = 200$, $\tau = 1$, and $\beta = 0.5$. We test the null hypothesis $H_{0, kink}^{\tau_0, \beta_0}$ with $\tau_0 = 1$ and $\beta_0 = 0.5$. The left panel plots $\hat{F}_{min}(y)$, the empirical CDF of $Y_{min}^{obs}$ (blue) and $\hat{F}_{max}^{tr}(y)$, the empirical CDF of the transformed outcomes $Y_{max}^{obs}$ (red). The test statistic $t_{KS}(1, 0.5)$ evaluated at the observed data is 0.07. The right panel plots the exact reference distribution of the test statistic given the sample and under the null hypothesis. The reference distribution is generated using 100,000 Monte Carlo draws from the $2^{200}$ randomization configurations. The dashed line plots the position of the observed test statistic. The exact $p$-value in this case is 0.61.
3.2. Testing the dose-response kink model. Let $H^K_0$ denote a composite hypothesis that is equal to the union of $H^{\tau_0,\beta_0}_{0\text{kink}}$ over all $\tau = \tau_0$ and $\beta = \beta_0$, i.e.,

$$H^K_0 = \bigcup_{\tau_0,\beta_0} H^{\tau_0,\beta_0}_{0\text{kink}},$$

where $(\tau, \beta)$ are nuisance parameters to be taken into account. One strategy testing $H^K_0$ is to take the supremum $p$-value over the entire range of $(\tau, \beta)$; another commonly used strategy is to first construct a confidence set around $(\tau, \beta)$ and then take the supremum $p$-values over the $(\tau, \beta)$ values in this confidence set (Berger and Boos, 1994). This latter strategy is particularly useful when the treatment dose and/or the outcome of interest are not bounded so that $\tau$ and $\beta$ are not bounded; see Nolen and Hudgens (2011), Ding, Feller and Miratrix (2016), and Zhang et al. (2021) for some applications of this strategy. In Supplementary Material C, we discuss how to construct a bounded level-$\gamma$ confidence set for $(\tau, \beta)$ based on inverting a variant of the Wilcoxon rank sum test statistic and its properties. Being able to reject $H^K_0$ suggests evidence against the postulated dose-response relationship; otherwise, the model is deemed sufficient to characterize the dose-response relationship for the data at hand.

We illustrate the procedure using the following example. We generate $I = 200$ matched pairs of 2 units with $Z_{ij}^{\text{obs}} \sim \text{Unif}[0, 4]$, $Y_{ij}(0) \sim \text{Normal}(0, 1)$, and $Y_{ij}^{\text{obs}} = Y_{ij}(Z_{ij}^{\text{obs}}) = Y_{ij}(0) + 2 \cdot 1\{0 \leq Z_{ij}^{\text{obs}} \leq 1\} + 1 \cdot 1\{1 < Z_{ij}^{\text{obs}} \leq 4\}$. Figure 4 plots the $p$-values in log scale against $\tau_0$ and $\beta_0$. The maximum $p$-value is obtained at $\tau_0 = 3.8$ and $\beta_0 = 0.4$ and equal to 0.004. The null hypothesis $H^K_0$, i.e., the dose-response relationship follows a kink model, can be rejected at level 0.05 for this simulated dataset.

**Fig 4**: The probability contour plot (in log scale) against values of $\tau_0$ and $\beta_0$. The true dose-response model is $Y_{ij}(z) = Y_{ij}(0) + 2 \cdot 1\{0 \leq z \leq 1\} + 1 \cdot 1\{1 < z \leq 4\}$. We let $Y_{ij}(0) \sim \text{N}(0, 1)$ and $I = 200$. We test $H^{\tau_0,\beta_0}_{0\text{kink}}$ and plot the $p$-value in log scale against $\tau_0$ and $\beta_0$ values. The maximum $p$-value is obtained at $\tau_0 = 3.8$ and $\beta_0 = 0.4$ and equal to 0.004. The null hypothesis $H^K_0$ is hence rejected at level 0.05 for this simulated dataset.
3.3. Testing any structured dose-response model. Our discussion above suggests a general model-free, randomization-based framework to test any structured dose-response relationship characterized by a few structural parameters. Consider the following structured dose-response relationship model where for all units:

$$H_0^{\text{dose-response}}: Y_{ij}(z) - Y_{ij}(z^*) \Delta f(z; z^*, \theta),$$

where $z^*$ is a reference dose, and $f(z; z^*, \theta)$ is a univariate function that satisfies $f(z^*; z^*, \theta) = 0$ and is parametrized by a $p$-dimensional vector of structural parameters $\theta \in \mathbb{R}^p$. Supplementary Material D.1 summarizes a general algorithm that tests $H_0^{\text{dose-response}}$ at level $\alpha$. We also briefly discuss and illustrate how to sequentially test a few dose-response relationships ordered in their model complexity in Supplementary Material D.2.

4. Dose-Response relationship under interference.

4.1. Potential outcomes under interference. We relax the stable unit treatment value assumption (SUTVA) in this section and consider inference for a structured dose-response relationship under interference. We collect the treatment doses of all study units in the matched-pair design in $Z = (Z_{11}, Z_{12}, \ldots, Z_{11}, Z_{12})$ with $z$ being its realization. Let $Z^{\text{obs}}$ be the observed treatment dose configuration of all $2I$ study units and

$$Y_{ij}(Z) := Y_{ij}(Z_{11}, \ldots, Z_{12})$$

be unit $ij$’s potential outcome that is random only via the randomness in $Z$. The SUTVA states that for all pairs of $z$ and $z'$, $z_{ij} = z'_{ij}$ implies $Y_{ij}(z) = Y_{ij}(z')$; in other words, $Y_{ij}(Z)$ depends on $Z$ only via its dependence on $Z_{ij}$.

Definition (6) is in a most general form and useful when the scientific interest lies in testing the null hypothesis of no direct or spillover effect under arbitrary interference pattern. To further explore the dose-response relationship in the presence of the spillover effect, researchers need to model the interference structure possibly based on units’ spatial relationship (e.g., closeness of counties in our case study). To this end, we assume study units are connected through an undirected network with a symmetric, $2I \times 2I$ adjacency matrix $G$. Matrix $G$ has its rows and columns arranged in the order corresponding to unit $11, 12, \ldots, I_1, I_2$ after nonbipartite matching. If unit $ij$ and $i'j'$ are connected, then the corresponding entry in $G$ is equal to 1 and otherwise 0. The diagonal entries of $G$ are defined to be 0.

Our reasoned basis for testing any causal null hypothesis under interference is still the approximate randomization scheme endowed by the nonbipartite matching. We have two goals. First, we show that the test developed for $H_{0, \text{null}}$ under the SUTVA remains a valid level-$\alpha$ test for a null hypothesis of no direct or spillover effect under an arbitrary interference pattern. Second, we relax the dose-response relationship $H_{0, \text{kink}}$ by modeling various forms of local interference pattern using the adjacency matrix $G$.

4.2. No direct or spillover effect. Following Rosenbaum (2007); Bowers, Fredrickson and Panagopoulos (2013); Athey, Eckles and Imbens (2018), a null hypothesis of no direct or spillover effect states that

$$H_{0, \text{direct or spillover}}: Y_{ij}(z) = Y_{ij}(z'),$$

for all units and treatment dose configurations $z$ and $z'$. Under $H_{0, \text{direct or spillover}}$, the unit-level potential outcome of each study unit under any treatment dose configuration $z$ can still be imputed; in fact, $Y_{ij}(z) = Y_{ij}(Z^{\text{obs}})$ for any $z$. Any test statistic (e.g., the Kolmogorov–Smirnov statistic) that depends on units’ potential outcomes (possibly under interference) is random only through its dependence on the treatment dose configurations of all study units; therefore,
the null distribution of the test statistic can again be inferred by enumerating configurations of $Z$, and the testing procedure for $H_{0, \text{null}}$ remains exact and has correct level for testing $H_{0, \text{direct or spillover}}$. Moreover, since $H_{0, \text{direct or spillover}}$ does not impose any interference pattern, rejecting $H_{0, \text{null}}$ implies rejecting $H_{0, \text{direct or spillover}}$ under arbitrary interference pattern.

4.3. Dose-Response relationship under local interference modeling. Testing the null hypothesis is often regarded a starting point of causal analysis (Imbens and Rubin, 2015). Next, we build up a causal hypothesis regarding a dose-response relationship allowing for local interference. Our construction is guided by the following three principles adapted from the literature on interference with a binary treatment (Hong and Raudenbush, 2006; Bowers, Fredrickson and Panagopoulos, 2013; Athey, Eckles and Imbens, 2018).

**Principle I:** The total effect of treatment dose configuration $z$ compared to a reference dose configuration $z^*$ can be decomposed into a dose-response direct effect due to $ij$’s own treatment dose $z_{ij}$ and a spillover effect due to other study units’ treatment doses so that $Y_{ij}(z) - Y_{ij}(z^*) = f(z_{ij}; z^*_{ij}, \theta) + g(z_{-ij}; z^*_{-ij})$ where $f(z_{ij}; z^*_{ij}, \theta)$ is a dose-response direct effect described in Section 3, $z_{-ij}$ (resp. $z^*_{-ij}$) treatment doses (resp. reference treatment doses) of all study units except $ij$, and $g(\cdot)$ a function modeling the spillover effect.

**Principle II:** The spillover effect depends only on the aggregate, excess treatment doses of $ij$’s neighbors with respect to the reference dose configuration so that $Y_{ij}(z) - Y_{ij}(z^*) = f(z_{ij}; z^*_{ij}, \theta) + g((z - z^*, G_{ij,\cdot}))$ where $G_{ij,\cdot}$ is the $ij$-th row of the adjacency matrix $G$.

**Principle III:** The spillover effect is always dominated by the dose-response direct effect in the sense that

$$\| G_{ij,\cdot} \|_0^{-1} \cdot (z - z^*, G_{ij,\cdot}) \leq z_{ij} - z^*_{ij} \quad \text{implies} \quad g((z - z^*, G_{ij,\cdot})) \leq f(z_{ij}; z^*_{ij}, \theta).$$

One simple modeling strategy of $g((z - z^*, G_{ij,\cdot}))$ that satisfies (7) is to scale the magnitude of the dose-response direct effect towards zero.

To illustrate the three principles above, we consider a concrete example of causal hypothesis under local interference. We consider a causal null hypothesis that states that the direct effect is proportional to the dose difference, i.e., $f(z_{ij}; z^*_{ij}, \theta) = \beta(z_{ij} - z^*_{ij})$. We then model the local interference pattern by scaling the direct effect using a logistic function so that $g((z - z^*, G_{ij,\cdot})) = C \times f(z_{ij}; z^*_{ij}, \theta)$ with $C = 1/(1 + \exp(-k((z - z^*, G_{ij,\cdot}) - s)))$. According to this specification, the spillover effect modeled by $g((z - z^*, G_{ij,\cdot}))$ trivially satisfies the third principle above as the multiplication factor $C$ is upper bounded by 1. The causal null hypothesis then becomes

$$H_{0, \text{interference}}: Y_{ij}(z) - Y_{ij}(z^*) = \beta(z_{ij} - z^*_{ij}) \cdot \left\{ \frac{1}{1 + \exp(-k((z - z^*, G_{ij,\cdot}) - s))} \right\}.$$ 

Statistical inference in the presence of interference parameters $(k, s)$ depends on one’s perspective on $(k, s)$ (Bowers, Fredrickson and Panagopoulos, 2013). Inference may proceed by regarding interference parameters as sensitivity parameters and researchers could report how confidence sets of the dose-response relationship parameters in the direct effect (e.g., $\beta$ in $H_{0, \text{interference}}$) change as interference parameters change. For fixed interference parameters $(k_0, s_0)$, we can test $\beta = \beta_0$ in $H_{0, \text{interference}}$ by imputing potential outcomes for each study unit under $H_{0, \text{interference}}$, generating the randomization-based reference distribution of a test statistic $t(Y(Z), Z)$, and comparing the observed test statistic to this reference distribution.

5. Social distancing and COVID-19 during phased reopening: study design.
5.1. Extension to longitudinal studies with a time-varying treatment dose. In our application, county-level social mobility evolved over time and public health outcomes could depend on the entire social mobility trajectory during the phased reopening period. To accommodate the longitudinal nature of our application, we extend the dose-response relationship to longitudinal studies with a period of treatment. We briefly summarize three major modeling assumptions below, and defer details to the Supplementary Material E.

A1. Sequential randomization assumption (SRA). In a longitudinal study, the observed outcome trajectory up to time \( t - 1 \) may confound the treatment dose at time \( t \); this is particularly true in our application: if the COVID-19 related case and death numbers were high during the last week in a county, then residents may be more wary of the disease and reduce social mobility this week. Following literature on longitudinal studies (Robins, 1998; Mattei, Ricciardi and Mealli, 2019), we adopt a version of the sequential randomization assumption (SRA) which states that residents’ social mobility at time \( t \) depends on (1) their history of adopting social distancing measures, (2) time-independent covariates, and (3) observed daily COVID-19 related case numbers and death toll up to time \( t - 1 \). We match on both the time-independent covariates (e.g., county demographics) and observed outcomes (e.g., observed case numbers and death toll) during the treatment period in the longitudinal setting.

A2. Cumulative dose assumption. We generalize the notion of dose to a time-varying treatment by defining a notion of cumulative dose. Let \( z_{t_0:t} \) be a realized treatment dose trajectory from time \( t_0 \) to \( t \). We define its cumulative dose to be a weighted sum of the difference between \( z_{t_0:t} \) and a reference trajectory \( z_{t_0:t}^* \), and denote it by \( CD(z_{t_0:t}; z_{t_0:t}^*, \mathcal{W}) \), where \( \mathcal{W} \) is a weighting scheme. The choices of the reference trajectory and weight function should be guided by expert knowledge so that the cumulative dose reflects some scientifically meaningful aspect of the treatment dose trajectory. For instance, in a longitudinal study of the effect of zidovudine (AZT), an antiretroviral medication, on mortality, Robins, Hernán and Babette (2000) defined the cumulative dose to be the aggregate AZT dose during the treatment period, i.e., the reference dose \( z_{t_0:t}^* = (0, \ldots, 0) \) and \( CD(z_{t_0:t}; z_{t_0:t}^*, \mathcal{W}) = \sum_{t_0 \leq t' \leq t} z_{t'} \). One major modeling assumption we adopt for the longitudinal setting is that the potential outcome depends on the time-varying treatment only via some cumulative dose. Although this assumption and its many variants are often assumed to reduce the number of potential outcomes (Robins, Hernán and Babette, 2000, Section 7), its validity needs to be evaluated on a case-by-case basis. We evaluated this assumption in the infectious disease dynamics context using standard compartment model before invoking it in our application; see Supplementary Material F for details.

A3. Time lag assumption. One important aspect of our application is that there is typically a time lag between social mobility and its effect on public health outcomes. To accommodate this, we further adopt a time lag assumption that says the outcome of interest at time \( t \) depends on the entire treatment dose trajectory only up to time \( t - \ell \). This assumption holds in particular when the outcome of interest is the number of people succumbing to the COVID-19 at time \( t \). Researchers estimated that the time lag between contracting COVID-19 and exhibiting symptoms (i.e., the so-called incubation period) had a median of 5.1 days and could be as long as 11.5 days (Lauer et al., 2020), and the time lag between the onset of the COVID-19 symptoms and death ranged from 2 to 8 weeks (Testa et al., 2020; World Health Organization, 2020). Therefore, it may be reasonable to believe that the number of COVID-19 related deaths at time \( t \) does not depend on social distancing practices \( \ell \) days immediately preceding \( t \) for some properly chosen \( \ell \).

Suppose that the time lag assumption holds for potential outcomes \( Y_{n,t+1}(-), \ldots, Y_{n,t+\ell}(-) \), and let \( g : \mathbb{R}^\ell \to \mathbb{R} \) be a function that maps these \( \ell \) potential outcomes to an aggregate outcome \( Y_{agg}(-) = g\{Y_{n,t+1}(-), \ldots, Y_{n,t+\ell}(-)\} \). One immediate consequence of
the time lag assumption is that \( Y_{\text{agg}}(\cdot) \) depends on \( Z_{t_0:t+\ell} \) only via \( Z_{t_0:t} \); moreover, we may further invoke the cumulative dose assumption so that \( Y_{\text{agg}}(\cdot) \) depends on \( Z_{t_0:t+\ell} \) only via some cumulative dose of \( Z_{t_0:t} \) by defining the cumulative dose with respect to a weighting scheme \( W_{\text{lag}} \) that assigns 0 weights to treatment doses corresponding to \( t = t+1, \ldots, t+\ell \).

5.2. Data: time frame, granularity, cumulative dose, outcome, and covariate history.

The first state in the U.S. that reopened was Georgia at April 24th, 2020. We hence consider data from April 27th, the first Monday following April 24th, to August 2nd, the first Sunday in August in the primary analysis. We choose a Monday (April 27th) as the baseline period and a Sunday (August 2nd) as the endpoint because social distancing and public health outcomes data exhibited consistent weekly periodicity (Unnikrishnan, 2020).

We analyze the data at a county-level granularity and use the county-level percentage change in the total distance traveled compiled by Unacast™ as the continuous, time-varying treatment dose. We consider a two-month treatment period from April 27th (Monday) to June 28th (Sunday). According to the data compiled by Unacast™, counties cut social mobility by at most 50% during most of the phased reopening; hence, we set the reference dose trajectory to be \(-0.5\) throughout the treatment period and define a notion of cumulative dose with respect to this reference dose trajectory and a uniform weighting scheme that assigns equal weights to each day during the treatment period. In a sensitivity analysis, we further repeated all dose-response analyses using different notions of cumulative dose based on different weighting schemes. The primary outcome of interest is the cumulative COVID-19 related death toll per 100,000 people from June 29th (Monday) to August 2nd (Sunday), a total of five weeks. The county-level COVID-19 case number and death toll are both obtained from the New York Times COVID-19 data repository (The New York Times, 2020).

We matched counties similar on covariates, including both time-independent covariates and time-dependent covariate histories. Specifically, we matched on the following time-independent baseline covariates: female (%), black (%), Hispanic (%), above 65 (%), smoking (%), driving alone to work (%), flu vaccination (%), some college (%), number of membership associations per 10,000 people, rural (0/1), poverty rate (%), population, and population density (residents per square mile). These county-level covariates data were derived from the census data collected by the United States Census Bureau and the County Health Rankings and Roadmaps Program (Remington, Catlin and Gennuso, 2015). Moreover, we matched on the number of new COVID-19 cases and new COVID-19 related deaths per 100,000 people every week from April 20th - 26th to June 23th - 29th.

5.3. Statistical matching, matched samples, and assessing balance.

A total of 1,211 matched pairs of two counties were formed using optimal nonbipartite matching (Lu et al., 2001, 2011). We matched exactly on the covariate “rural (0/1)” for later subgroup analysis and balanced all other 32 covariates. We added a mild penalty on the cumulative dose so that two counties within the same matched pair had a tangible difference in their cumulative doses, and added 20% sinks to eliminate 20% of counties for which no good match can be found (Baiocchi et al., 2010). Following Rubin (2007), the design was conducted without any access to the outcome data to assure the objectivity of the design.

Within each matched pair, the county with smaller cumulative dose is referred to as the “better social distancing” county, and the other the “worse social distancing” county. Appendix A shows where the 1,211 better social distancing counties and the other 1,211 worse social distancing counties are located in the U.S., and Figure 5 plots the average daily percentage change in total distance traveled and the average daily COVID-19 related death toll per 100,000 people during the treatment period (April 27th to June 28th) in two groups. It is evident that two groups differ in their extent of social distancing, but are very similar in their
daily COVID-19 related death toll per 100,000 people during the treatment period. Finally, Appendix B summarizes the balance of all 33 covariates in two groups after matching. All variables have standardized mean differences less than 0.15 and are considered sufficiently balanced (Rosenbaum, 2002). In Supplementary Material G.1, we further plot the cumulative distribution functions of important variables in two groups. A detailed pre-analysis plan, including matched samples and specification of a primary analysis and three secondary analyses, can be found via doi:10.13140/RG.2.2.23724.28800.

![Graph](image)

**Fig 5:** Trajectories of the average daily percentage change in total distance traveled (dashed lines) and average daily COVID-19 related death toll per 100,000 people (solid lines) in 1,211 better social distancing counties (blue) and 1,211 worse social distancing counties (red). We saw a sharp contrast in the mobility but little difference in the COVID-19 related death toll during this treatment period.

### 6. Social distancing and COVID-19 during phased reopening: outcome analysis.

#### 6.1. Primary analysis: causal null hypothesis regarding the death toll

Fix \( t_0 = \) April 27th and \( T = \) June 28th. We specify the time-lag parameter \( \ell = 35 \) so that \( T + \ell \) corresponds to August 2nd. Let \( Y_t(\cdot) \) be the potential COVID-19 related deaths at time \( t \), and consider the aggregate outcome \( Y_{agg}(\cdot) = \sum_{T+1 \leq t \leq T+\ell} Y_t(\cdot) \). Our primary analysis tests the causal null hypothesis for all counties in our matched samples:

\[
H_{0, primary}: Y_{ij, agg}(z_{t_0:T}) - Y_{ij, agg}(z_{t_0:T}) = 0,
\]

for all \( i, j \), and \( z_{t_0:T} \). This null hypothesis states that the treatment dose trajectory from April 27th to June 28th had no effect whatsoever on the COVID-19 related death toll from June 29th to August 2nd.

The top left panel of Figure 6 plots the empirical distribution functions of the better and worse social distancing counties’ observed outcomes; we calculate the test statistic \( t_{KS} = 0.735 \) and contrast it to a reference distribution generated using 1,000,000 samples.
from all possible $2^{1211}$ randomizations; see the top right panel of Figure 6. In this way, an exact $p$-value $< 0.001$ is obtained and the causal null hypothesis $H_{0,\text{primary}}$ is rejected at 0.05 level. Moreover, as detailed in Section 4, rejecting $H_{0,\text{primary}}$ also implies rejecting the null hypothesis of no direct or spillover effect under arbitrary interference pattern.

We further conducted two sensitivity analyses to assess sensitivity to the no unmeasured confounding assumption and the time lag assumption in the primary analysis. In the first sensitivity analysis, we allowed the dose trajectory assignment probability $\pi_{i1}$ and $\pi_{i2}$ to be biased from the randomization probability and then generated the reference distribution with this biased randomization probability; specifically, we considered a biased treatment assignment model where $\log(\Gamma_i) = \log(\pi_{i1}/\pi_{i2})$ was proportional to the absolute difference in the cumulative doses of two units in each pair ($\pi_{i1} = \pi_{i2} = 1/2$ and $\Gamma_i = 1$ in a randomized experiment for all $i$). We found that our primary analysis conclusion would hold up to $\Gamma_i$ having a median as large as 3.82. See Supplementary Material G.3.1 for details. In the second sensitivity analysis, we repeated the primary analysis using a shorter time lag $l = 28$ days and the result was similar; see Supplementary Material G.3.2 for details.

Our primary analysis results suggest that different social mobility trajectories during the treatment period had an effect on the COVID-19-related death toll in the subsequent weeks. This conclusion stands under an arbitrary interference and is robust to unmeasured confounding.

---

**Fig 6:** Top left panel: empirical distribution functions of cumulative COVID-19 related deaths per 100,000 people in the better social distancing (blue) and worse social distancing (red) groups. Top right panel: randomization-based reference distribution. The exact $p$-value is $< 0.001$. Bottom left panel: empirical distribution functions of cumulative COVID-19 cases per 100,000 people in the better social distancing (blue) and worse social distancing (red) groups. Bottom right panel: randomization-based reference distribution. The exact $p$-value is $< 0.001$. 

6.2. **Secondary analysis I: secondary outcome.** Let \( Y_{ij,\text{case,agg}} \) denote the cumulative COVID-19 cases per 100,000 people from June 29th to July 12th (corresponding to time lag \( l = 14 \) days), as specified in our pre-analysis plan. We test the following null hypothesis concerning the secondary outcome \( Y_{ij,\text{case,agg}} \):

\[
H_0,\text{secondary} : \quad Y_{ij,\text{case,agg}}(z_{t0:T}) - Y_{ij,\text{case,agg}}(z_{t0:T}^*) = 0,
\]

for all \( i, j \), and \( z_{t0:T} \). The exact \( p \)-value is \(< 0.001\); see the bottom panels of Figure 6. In a sensitivity analysis, we repeated the test with a shorter time lag \( l = 10 \) days and the result was similar; see Supplementary Material G.3.3 for details. Our result suggests strong evidence that social distancing from April 27th to June 28th had an effect on cumulative COVID-19 cases per 100,000 people from June 29th to July 12th in our matched samples.

6.3. **Secondary analysis II: dose-response relationship.** Let \( z_{t0:T}^* \) be a reference trajectory equal to the \(-0.50\) for all \( t_0 \leq t \leq T \) (corresponding to 50\% reduction in total distance traveled from April 27th to June 28th), \( W_{\text{lag}} \) a weighting scheme that assigns uniform weights to all \( t \) such that \( t_0 \leq t \leq T \) and 0 otherwise, and \( \text{CD}(z_{t0:T}; z_{t0:T}^*, W_{\text{lag}}) \) a cumulative dose. We invoke the cumulative dose assumption A2 so that \( Y_{ij,\text{case,agg}}(\cdot) \) depends on \( z_{t0:T} \) only via \( \text{CD}(z_{t0:T}; z_{t0:T}^*, W_{\text{lag}}) \), and test the following dose-response kink model for all units:

\[
H_{0,\text{kink}} : \quad Y_{ij,\text{case,agg}}(z_{t0:T}) - Y_{ij,\text{case,agg}}(z_{t0:T}^*) = 0, \quad z_{t0:T} \text{ such that } \text{CD}(z_{t0:T}; z_{t0:T}^*, W_{\text{lag}}) \leq \tau,
\]

and

\[
\log \{Y_{ij,\text{case,agg}}(z_{t0:T})\} - \log \{Y_{ij,\text{case,agg}}(z_{t0:T}^*)\} = \beta \cdot \{\text{CD}(z_{t0:T}; z_{t0:T}^*, W_{\text{lag}}) - \tau\},
\]

\( z_{t0:T} \) such that \( \text{CD}(z_{t0:T}; z_{t0:T}^*, W_{\text{lag}}) > \tau \).

This dose-response relationship states that the potential COVID-19 cases from June 29th to July 12th remains unchanged as the potential outcome under \( Z_{t0:T} = z_{t0:T}^* \), i.e., strict social distancing that reduces total distance traveled by 50\% everyday from April 27th to June 28th, when the cumulative dose is less than some threshold \( \tau \); after the cumulative dose exceeds this threshold, the COVID-19 case number increases exponentially at a rate proportional to how much the cumulative dose exceeds the threshold.

---

**Fig 7:** Left panel: contour plot of \( p \)-values when testing \( H_{0,\text{kink}} \) as in (8) against \( \tau = \tau_0 \) and \( \beta = \beta_0 \). Maximum \( p \)-value is obtained at \((\tau_0, \beta_0) = (0.48, 10.0)\) (red marker). Three isopleths (0.1, 0.05, and 0.005) are plotted. Right panel: dose-response relationships for selected \((\tau_0, \beta_0)\) in the 0.1 confidence set with baseline \( Y_{ij,\text{case,agg}}(z_{t0:T}^*) \) equal to 1 per 100,000. The red line corresponds to \((\tau_0, \beta_0) = (0.46, 5.0)\), blue line \((\tau_0, \beta_0) = (0.48, 10.0)\), and orange line \((\tau_0, \beta_0) = (0.50, 12.0)\).
We tested $H_{0,kink}$ for different $\tau = \tau_0$ and $\beta = \beta_0$ combinations; the maximum $p$-value was obtained at $(\tau_0, \beta_0) = (0.48, 10.0)$ and equal to 0.417; hence, the kink model was not rejected. The left panel of Figure 7 plots the level-0.1 and level-0.05 confidence sets of $(\tau, \beta)$. The right panel of Figure 7 plots three dose-response curves with baseline (i.e., 50% reduction) case number equal to 1 per 100,000 people for three selected $(\tau_0, \beta_0)$ pairs in the level-0.1 confidence set. In a sensitivity analysis, we repeated the analysis by considering two different specifications of cumulative dose, one assigning more weight to early days of the phased reopening and the other towards the end of the phased reopening. Confidence sets look similar under different specifications; see Supplementary Material G.3.4 for details.

The confidence set of the threshold parameter $\tau$ is tightly centered around 0.5, suggesting that as a county’s average percentage change in total distance traveled from April 27th to June 28th increases from $-50\%$ to around $-5\%$ to $5\%$, the potential COVID-19 cases number from June 29th to July 12th would largely remain unchanged; however, once beyond this threshold, the case number would rise exponentially and could incur an increase as large as 10-fold when a county’s average distance traveled increased by about $20\%$ compared to the pre-coronavirus level.

6.4. Secondary analysis III: subgroup analysis and differential dose-response relationship. We also conducted subgroup analyses by repeating the primary and secondary analyses described in Section 6.1 to 6.3 on 462 matched pairs of 2 non-rural counties and 749 matched pairs of 2 rural counties. $P$-values testing the primary analysis hypothesis concerning the death toll and secondary analysis hypothesis concerning the case numbers are 0.004 and < 0.001, respectively, in the non-rural subgroup, and 0.008 and 0.009, respectively, in the rural subgroup; see Supplementary Material G.2 for details.

We also allowed a differential dose-response relationship between mobility and case numbers in rural and non-rural counties and constructed confidence sets for $(\tau, \beta)$ separately; see Figure 8. We further repeated the subgroup analyses under different specifications of the cumulative dose and the results were similar; see Supplementary Material G.3.4 for details.

A comparison of confidence sets for the non-rural (top left panel of Figure 8) and rural counties (bottom left panel of Figure 8) revealed an intriguing pattern: while the level-0.1 confidence set of the activation threshold $\tau$ is centered around 0.5 for the rural counties, it contains 0 for the non-rural counties; moreover, the level-0.1 confidence set of rural counties covers a much larger range of $\beta$ values compared to that of the non-rural counties. Together, these results suggest that the activation dose required to trigger exponential growth in case numbers for rural counties seemed larger than that for non-rural counties; however, once exponential growth was incurred, the growth seemed more rapid in rural counties.

6.5. Dose-Response relationship under local interference. We next construct corrected confidence sets of $(\tau, \beta)$ under local interference modeling. Let $(z_{i_0:T})_{i,j} = (z_{i_1:t_{i_0:T}}, \ldots, z_{i_2:t_{i_0:T}})$ be an ordered set of treatment dose trajectories of all study units, and $(z_{t_0:T})_{i,j}$ an ordered set of reference dose trajectories. We collect the cumulative doses of all study units during the treatment period in $z_{\text{cumu}} = (\text{CD}(z_{i_1:t_{i_0:T}}; z_{i_0:T}; W_{\text{lag}}), \ldots, \text{CD}(z_{i_2:t_{i_0:T}}; z_{i_0:T}; W_{\text{lag}}))$, and
Fig 8: Top left panel: contour plot of $p$-values when testing $H_{0, \text{kink}}$ against $\tau = \tau_0$ and $\beta = \beta_0$ in 462 matched pairs of 2 non-rural counties. Three isopleths (0.1, 0.05, and 0.005) are plotted. Top right panel: dose-response relationships for selected $(\tau_0, \beta_0)$ in the 0.1 confidence set as in the top left panel with baseline $Y_{ij, \text{case, agg}}(z_{t_0:T}^*)$ equal to 1 per 100,000. The red line corresponds to $(\tau_0, \beta_0) = (0.10, 3.0)$, blue line $(\tau_0, \beta_0) = (0.38, 8.0)$, and orange line $(\tau_0, \beta_0) = (0.41, 12.0)$. Bottom left panel: contour plot of $p$-values when testing $H_{0, \text{kink}}$ against $\tau = \tau_0$ and $\beta = \beta_0$ in 749 matched pairs of rural counties. Three isopleths (0.1, 0.05, and 0.005) are plotted. Bottom right panel: dose-response relationships for selected $(\tau_0, \beta_0)$ in the 0.1 confidence set as in the bottom left panel with baseline $Y_{ij, \text{case, agg}}(z_{t_0:T}^*)$ equal to 1 per 100,000. The red line corresponds to $(\tau_0, \beta_0) = (0.5, 10.0)$, blue line $(\tau_0, \beta_0) = (0.55, 15.0)$, and orange line $(\tau_0, \beta_0) = (0.60, 20.0)$.

incorporate local interference as follows:

$$H_{0, \text{kink, inter}} : Y_{ij, \text{case, agg}}((z_{t_0:T})_{i,j}) - Y_{ij, \text{case, agg}}((z_{t_0:T}^*)_{i,j}) = 0,$$

such that $CD(z_{ij; t_0:T}; z_{t_0:T}^*, W_{lag}) \leq \tau$,

and

$$\log(Y_{ij, \text{case, agg}}(z_{t_0:T})) - \log(Y_{ij, \text{case, agg}}(z_{t_0:T}^*)) = \beta \cdot \{CD(z_{ij; t_0:T}; z_{t_0:T}^*, W_{lag}) - \tau\} \cdot \left\{1 + \frac{1}{1 + \exp\{-k(\|G_{ij, s}\|_0^{-1} \cdot (z_{\text{cumu}}, G_{ij, s}) - s)\}\cdot \text{Spillover Effect Factor}} \right\},$$

such that $CD(z_{ij; t_0:T}; z_{t_0:T}^*, W_{lag}) > \tau$.

According to this null hypothesis, there is no direct effect if county $ij$’s cumulative dose is below some threshold $\tau$; hence, there is no spillover effect in this case by Principle III described in Section 4. Once county $ij$’s cumulative dose is above the threshold, this triggers exponential growth captured by the dose-response direct effect plus a spillover effect. The magnitude of the spillover effect is equal to the direct effect multiplied by a spillover effect.
factor $C$. This spillover effect factor depends on the average cumulative dose of $ij$’s neighbors but is always upper bounded by 1 so that the spillover effect is no larger than the direct effect (see Section 4.3). In rare cases when a county has no neighbor, $C$ is defined to be 0 so that there is no spillover effect. We used the county adjacency file provided by the United States Census Bureau (U.S Census Bureau) as our adjacency matrix $G$.

The interference parameters $(k, s)$ in the above model are easy to interpret and specify. For instance, $(k, s) = (5, 1)$ corresponds to a small spillover effect of approximately 1% of the direct effect when neighbors’ average cumulative dose is 0.10 (corresponding to an average 40% reduction in social mobility compared to the pre-pandemic level) and approximately 8% of the direct effect when neighbors’ average cumulative dose is 0.50 (corresponding to social mobility remaining the same as the pre-pandemic level). In this way, the parameters $(k, s)$ carry concrete meanings and can be easily tuned and communicated to the audience.

The left panel of Figure 9 plots the level-0.1, 0.05, and 0.005 confidence sets of $(\tau, \beta)$ when $(k, s) = (5, 1)$. The right panel of Figure 9 further illustrates the inferred dose-response direct effects (solid lines) and the corresponding spillover effects (dotted lines) under $(k, s) = (5, 1)$ for $(\tau, \beta) = (0.44, 2.5)$ (red lines) and $(0.48, 5.0)$ (blue lines).

The level-0.05 confidence set of the dose-response direct effect contains similar $\tau$ values but considerably smaller $\beta$ values compared to assuming no interference and modeling the total effect using a dose-response kink model (see the left panel of Figure 7). This makes intuitive sense as the total effect has now been decomposed into the dose-response direct effect and a spillover effect due to neighboring counties.

![Figure 9](image_url)

**Fig 9:** Left panel: contour plot of $p$-values when testing $H_{0,kink,inter}$ against $\tau = \tau_0$ and $\beta = \beta_0$ under interference parameters $(k, s) = (5, 1)$. Maximum $p$-value is obtained at $(\tau_0, \beta_0) = (0.44, 2.5)$ (red marker). Three isopleths (0.1, 0.05, and 0.005) are plotted. Right panel: dose-response direct effect (solid lines) and the associated spillover effects (dotted lines) for selected $(\tau_0, \beta_0)$ in the 0.05 confidence set with baseline $Y_{ij,case,agg}(z_{ij,T}^\tau)$ equal to 1 per 100,000. Two red lines correspond to $(\tau_0, \beta_0) = (0.44, 2.5)$ and blue lines $(\tau_0, \beta_0) = (0.48, 5.0)$.

7. Discussion. We studied in detail the effect of social distancing during the early phased reopening in the United States on COVID-19 related death toll and case numbers using our compiled county-level data. To address the statistical challenge brought by a time-dependent, continuous treatment dose trajectory, we developed a design-based framework based on non-bipartite matching to embed observational data with time-dependent, continuous treatment dose trajectory into a randomized controlled experiment. This embedding induces a randomization scheme that we then used to conduct randomization-based, model-free statistical inference for causal relationships, including testing a causal null hypothesis, a structured dose-response relationship and a causal null hypothesis under local interference modeling.
Upon applying the proposed design and testing procedures to the mobility and COVID-19 data, we found very strong evidence against the causal null hypothesis and supportive of a causal effect of social distancing during the early phases of reopening on subsequent COVID-19-related death and case numbers. Our finding complements many recent studies based on standard epidemiological models (Dickens et al., 2020; Koo et al., 2020) and structural equation modeling (Chernozhukov, Kasahara and Schrimpf, 2021; Bonvini et al., 2021) from a unique perspective, and once again confirms the important role of social distancing (as captured by a reduction in mobility in this article) in combating the novel coronavirus. Our transparent comparison of two groups of similar counties makes our findings digestible and easy to communicate to the general public.

In a dose-response analysis, we found that the confidence set of the dose needed to activate exponential growth was tightly centered and its magnitude suggested that once the total distance traveled returned to or even superseded the pre-coronavirus level, it would have a devastating effect on the COVID-19 case numbers by contributing to exponential growth. Moreover, in a subgroup analysis where we allowed a differential dose-response relationship, we found that more stringent social distancing would be needed to avoid devastating exponential growth for non-rural counties; however, once the exponential growth was incurred, the growth appeared more rapid in rural counties. This striking difference in dose-response relationship between rural and non-rural communities agrees with experts’ assessment of the transmission dynamics. Given its clinical features, the rate of virus reproduction is likely higher in large, urban areas due to more reproductive opportunities afforded by denser populations (Souch and Cossman, 2020) and this may explain the absence of an “activation dose” in non-rural counties (see top left panel of Figure 8). On the other hand, although rural residents have less social interaction compared to non-rural counterparts, they often have more underlying medical conditions and are more likely to present for treatment at more advanced stages of disease (Callaghan et al., 2021), which may partly explain why rural communities seemed to incur more drastic exponential growth in case numbers once the activation dose was exceeded (see bottom left panel of Figure 8).

The design-based approach and analysis proposed in this article has its limitations. First, we used social mobility data as a proxy measure for social distancing. It would be interesting to look at other aspects of social distancing, e.g., closure of borders, reduction in aviation travel, etc, in future works. Second, in order to permute two treatment dose trajectories in a longitudinal setting, one necessarily needs to match on observed outcomes during the treatment period and compare outcomes after the treatment period; therefore, in a longitudinal setting, the method is suited only for applications where the effect of a time-varying treatment is not immediate, e.g., effect of precautionary measures on the death toll. Third, when the sample size is limited, the interference parameters are treated as sensitivity parameters that researchers vary, rather than population parameters for which researchers construct confidence sets. The proposed method also has strengths: it embeds the noisy observational data into an approximate randomized controlled trial and has a clear “reasoned basis” (Fisher, 1935) when testing the causal null hypothesis, and researchers can always conduct a sensitivity analysis to investigate how causal conclusions would change when the randomization assumption is relaxed. The method developed in this article can be readily applied to many practical problems where there is a continuous exposure and the scientific interest lies in testing a dose-response relationship. Understanding a dose-response relationship is central to many scientific disciplines like public health (Gorell et al., 1999; Farrelly et al., 2005), pharmacology (Tallarida and Jacob, 2012), and toxicology (Calabrese and Baldwin, 2003), among many others.
Fig 10: Map of 1,211 better social distancing (light blue) and 1,211 worse social distancing counties (red) in the matched analysis. Unmatched counties are in white.

APPENDIX B: BALANCE TABLE AFTER STATISTICAL MATCHING

SUPPLEMENTARY MATERIAL

Pilot study, technical details, and further details on the case study
Supplementary Material A provides details on the pilot study described in Section 1.1 in the main article. Supplementary Material B motivates the Kolmogorov–Smirnov-type test statistic considered in the main article. Supplementary Material C discusses how to construct a confidence set for nuisance parameters $\tau, \beta$ in a dose-response kink model based on a variant of the rank sum test. Supplementary Material D summarizes how to test a structural dose-response relationship and illustrates how to test a sequence of dose-response models ordered according to their model complexity. Supplementary Material E extends the design to longitudinal studies and elaborates on necessary modeling assumptions. Supplementary Material F assesses the cumulative dose assumption using a standard epidemiological model. Supplementary Material G provides further details on the case study, including a closer examination of the distributions of some important variables after matching, separate analyses of rural and non-rural counties, and numerous sensitivity analyses.

code and data.zip
Data and R code implementing the statistical matching and randomization inference.
**Table 3**

Covariate balance after statistical matching. A total of 1,211 pairs of two counties are formed. All absolute standardized mean differences are less than 0.15.

| Time-Independent Covariates | Better Social Distancing Counties | Worse Social Distancing Counties | Standardized Difference |
|-----------------------------|-----------------------------------|----------------------------------|--------------------------|
|                            | \( n = 1,211 \)                    | \( n = 1,211 \)                  |                          |
| Female (fr)                 | 0.50                              | 0.50                             | 0.06                     |
| Above 65 (fr)               | 0.20                              | 0.19                             | -0.05                    |
| Black (fr)                  | 0.07                              | 0.07                             | 0.00                     |
| Hispanic (fr)               | 0.09                              | 0.08                             | -0.03                    |
| Driving alone to work (fr)  | 0.80                              | 0.81                             | 0.12                     |
| Smoking (fr)                | 0.17                              | 0.18                             | 0.15                     |
| Flu vaccination (fr)        | 0.42                              | 0.42                             | -0.01                    |
| Some college (fr)           | 0.59                              | 0.58                             | -0.09                    |
| Membership association (per 10,000 people) | 12.29 | 12.01 | -0.05 |
| Rural (0/1)                 | 0.62                              | 0.62                             | 0.00                     |
| Below poverty (fr)          | 0.14                              | 0.15                             | 0.14                     |
| Population density (residents per \( \text{mi}^2 \)) | 173 | 130 | -0.08 |
| Population                  | 92,310                            | 79,423                           | -0.06                    |

| Time-Varying Covariates     |                                  |                                  |                          |
|-----------------------------|----------------------------------|----------------------------------|--------------------------|
| Cases per 100,000 people    |                                  |                                  |                          |
| Apr 20th - Apr 26th         | 27.97                            | 25.30                            | -0.02                    |
| Apr 27th - May 3rd          | 29.74                            | 24.23                            | -0.08                    |
| May 4th - May 10th          | 29.17                            | 25.25                            | -0.05                    |
| May 11th - May 17th         | 26.21                            | 25.49                            | -0.01                    |
| May 18th - May 24th         | 29.13                            | 28.70                            | -0.01                    |
| May 25th - June 1st         | 30.95                            | 25.62                            | -0.07                    |
| June 2nd - June 8th        | 29.75                            | 28.78                            | -0.01                    |
| June 9th - June 15th       | 28.40                            | 31.31                            | 0.04                     |
| June 16th - June 22th      | 34.02                            | 40.97                            | 0.09                     |
| June 23th - June 29th      | 45.51                            | 51.94                            | 0.09                     |
| Deaths per 100,000 people  |                                  |                                  |                          |
| Apr 20th - Apr 26th        | 1.35                             | 0.92                             | -0.12                    |
| Apr 27th - May 3rd         | 1.20                             | 0.95                             | -0.08                    |
| May 4th - May 10th         | 1.35                             | 1.00                             | -0.09                    |
| May 11th - May 17th        | 1.00                             | 0.98                             | -0.01                    |
| May 18th - May 24th        | 0.95                             | 0.91                             | -0.01                    |
| May 25th - June 1st        | 1.06                             | 0.84                             | -0.07                    |
| June 2nd - June 8th        | 0.85                             | 0.70                             | -0.06                    |
| June 9th - June 15th       | 0.67                             | 0.66                             | -0.00                    |
| June 16th - June 22th      | 0.62                             | 0.64                             | 0.01                     |
| June 23th - June 29th      | 0.85                             | 0.68                             | -0.05                    |

**References**

BBC Radio 4 (2020). Best of Today. [https://www.bbc.co.uk/programmes/p08jn7g4](https://www.bbc.co.uk/programmes/p08jn7g4).

Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M. D. (2020). A multi-risk SIR model with optimally targeted lockdown Technical Report, National Bureau of Economic Research.

Atalan, A. (2020). Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective. *Annals of Medicine and Surgery* 56 38–42.

Athey, S., Eckles, D. and Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association* 113 230–240.

Baiocchi, M., Small, D. S., Lorch, S. and Rosenbaum, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association* 105 1285–1296.

Berger, R. L. and Boos, D. D. (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association* 89 1012–1016.

Bind, M.-A. C. and Rubin, D. B. (2019). Bridging observational studies and randomized experiments by embedding the former in the latter. *Statistical Methods in Medical Research* 28 1958–1978.
Bonvini, M., Kennedy, E., Ventura, V. and Wasserman, L. (2021). Causal inference in the time of Covid-19. arXiv preprint arXiv:2103.04472.

Bowers, J., Fredrickson, M. M. and Panagopoulos, C. (2013). Reasoning about interference between units: A general framework. Political Analysis 97–124.

Brauer, F. and Castillo-Chavez, C. (2012). Mathematical Models in Population Biology and Epidemiology. Springer.

U.S. Census Bureau County Adjacency File.

Calabrese, E. J. and Baldwin, L. A. (2003). Hormesis: the dose–response revolution. Annual Review of Pharmacology and Toxicology 43 175–197.

Callaghan, T., Lueck, J. A., Trujillo, K. L. and Ferdinand, A. O. (2021). Rural and urban differences in COVID-19 prevention behaviors. The Journal of Rural Health 37 287–295.

Chernozhukov, V., Kasahara, H. and Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the US. Journal of Econometrics 220 23–62.

Dickers, B. L., Koo, J. R., Wilder-Smith, A. and Cook, A. R. (2020). Institutional, not home-based, isolation could contain the COVID-19 outbreak. The Lancet 395 1541–1542.

Ding, P., Feller, A. and Miratrix, L. (2016). Randomization inference for treatment effect variation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78 655–671.

Farrelly, M. C., Davis, K. C., Haviland, M. L., Messeri, P. and Healton, C. G. (2005). Evidence of a dose–response relationship between “truth” antismoking Ads and youth smoking prevalence. American Journal of Public Health 95 425–431.

Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd. London and Edinburgh.

Gelfand, M. J., Jackson, J. C., Pan, X., Nau, D., Pieper, D., Denison, E., Dagher, M., Van Lange, P. A., Chiu, C.-Y. and Wang, M. (2021). The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. The Lancet Planetary Health 5 e135–e144.

Gorell, J. M., Rybicki, B. A., Johnson, C. C. and Peterson, E. L. (1999). Smoking and Parkinson’s disease: a dose–response relationship. Neurology 52 115–115.

Grover, S., Sahoo, S., Mehra, A., Avasthi, A., Tripathi, A., Subramanyan, A., Pattooiashi, A., Rao, G. P., Saha, G., Mishra, K. et al. (2020). Psychological impact of COVID-19 lockdown: An online survey from India. Indian Journal of Psychiatry 62 354.

Heng, S., Zhang, B., Han, X., Lorch, S. A. and Small, D. S. (2019). Instrumental Variables: to Strengthen or not to Strengthen? arXiv preprint arXiv:1911.09171.

Ho, D. E., Imai, K., King, G. and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15 199–236.

Hong, G. and Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. Journal of the American Statistical Association 101 901–910.

Imbens, G. W. and Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Koo, J. R., Cook, A. R., Park, M., Sun, Y., Sun, H., Lim, J. T., Tam, C. and Dickens, B. L. (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases 20 678–688.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J. and Khosrawipur, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. Journal of Travel Medicine 27 taaa037.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G. and Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of Internal Medicine 172 577–582.

Lewnard, J. A. and Lo, N. C. (2020). Scientific and ethical basis for social-distancing interventions against COVID-19. The Lancet. Infectious diseases 20 631.

Lu, B., Zanutto, E., Hornik, R. and Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. Journal of the American Statistical Association 96 1245–1253.

Lu, B., Greely, R., Xu, X. and Beck, C. (2011). Optimal nonbipartite matching and its statistical applications. The American Statistician 65 21–30.

Mattei, A., Ricciardi, F. and Mealli, F. (2019). Bayesian Inference for Sequential Treatments Under Latent Sequential Ignorability. Journal of the American Statistical Association.

Nolen, T. L. and Hudgens, M. G. (2011). Randomization-based inference within principal strata. Journal of the American Statistical Association 106 581–593.

World Health Organization (2020). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19).

Remington, P. L., Catlin, B. B. and Gennuso, K. P. (2015). The county health rankings: rationale and methods. Population Health Metrics 13 11.
ROBINS, J. M. (1998). Marginal structural models. In: 1997 Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association, 1998;1-10.

ROBINS, J., HERNÁN, M. A. and BARETTE, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology 11 550-560.

ROSENBAUM, P. R. (1989). Sensitivity analysis for matched observational studies with many ordered treatments. Scandinavian Journal of Statistics 227–236.

ROSENBAUM, P. R. (2002). Observational Studies. Springer.

ROSENBAUM, P. R. (2005). Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. The American Statistician 59 147–152.

ROSENBAUM, P. R. (2007). Interference between units in randomized experiments. Journal of the American Statistical Association 102 191–200.

ROSENBAUM, P. R. (2010). Design of Observational Studies. Springer.

RUBIN, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. Journal of the American Statistical Association 75 591–593.

RUBIN, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. Journal of the American Statistical Association 81 961–962.

RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association 100 322–331.

RUBIN, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in Medicine 26 20–36.

SHERIDAN, A., ANDERSEN, A. L., HANSEN, E. T. and JOHANNESEN, N. (2020). Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. Proceedings of the National Academy of Sciences 117 20468–20473.

SJÖDIN, H., WILDER-SMITH, A., OSMAN, S., FAROOQ, Z. and ROCKLÖV, J. (2020). Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Eurosurveillance 25 2000280.

SOUCH, J. M. and COSSMAN, J. S. (2020). A commentary on rural–urban disparities in COVID-19 testing rates per 100,000 and risk factors. The Journal of Rural Health.

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science 25 1–21.

TALLARIDA, R. J. and JACOB, L. S. (2012). The dose—response relation in pharmacology. Springer Science & Business Media.

TESTA, C. C., KRIEGER, N., CHEN, J. T. and HANAGE, W. P. (2020). Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020). Technical Report, The Harvard Center for Population and Development Studies, Munich.

THE NEW YORK TIMES (2020). Coronavirus (Covid-19) Data in the United States. https://github.com/nytimes/covid-19-data. Accessed: 2020-09-30.

UNNIKRISHNAN, C. (2020). Globally Coherent Weekly Periodicity in the Covid-19 Pandemic. medRxiv.

VENKATESH, A. and EDIRAPPULI, S. (2020). Social distancing in Covid-19: what are the mental health implications? BMJ 369.

ZHANG, B., HENG, S., MACKAY, E. J. and YE, T. (2021). Bridging preference-based instrumental variable studies and cluster-randomized encouragement experiments: Study design, noncompliance, and average cluster effect ratio. Biometrics in press.