Assessing the Causal Impact of COVID-19 Related Policies on Outbreak Dynamics: A Case Study in the US

Jing Ma[1], Yushun Dong[1], Zheng Huang[1], Daniel Mietchen[1,2], Jundong Li[1]

[1]University of Virginia, Charlottesville, VA, USA 22904
[2]Fraunhofer Institute for Biomedical Engineering, Sulzbach, Germany
{jm3mr, yd6eb, zh4zn, jundong}@virginia.edu, daniel.mietchen@ibmt.fraunhofer.de

ABSTRACT

Analyzing the causal impact of different policies in reducing the spread of COVID-19 is of critical importance. The main challenge here is the existence of unobserved confounders (e.g., vigilance of residents) which influence both the presence of policies and the spread of COVID-19. Besides, as the confounders may be time-varying, it is even more difficult to capture them. Fortunately, the increasing prevalence of web data from various online applications provides an important resource of time-varying observational data, and enhances the opportunity to capture the confounders from them, e.g., the vigilance of residents over time can be reflected by the popularity of Google searches about COVID-19 at different time periods. In this paper, we study the problem of assessing the causal effects of different COVID-19 related policies on the outbreak dynamics in different counties at any given time period. To this end, we integrate COVID-19 related observational data covering different U.S. counties over time, and then develop a neural network based causal effect estimation framework which learns the representations of time-varying (unobserved) confounders from the observational data. Experimental results indicate the effectiveness of our proposed framework in quantifying the causal impact of policies at different granularities, ranging from a category of policies with a certain goal to a specific policy type. Compared with baseline methods, our assessment of policies is more consistent with existing epidemiological studies of COVID-19. Besides, our assessment also provides insights for future policy-making.

CCS CONCEPTS

• **Applied computing** → **Law, social and behavioral sciences**;
• **Mathematics of computing** → *Causal networks*.

KEYWORDS

COVID-19, Causal Inference, Individual Treatment Effect, Network, Observational data

1 INTRODUCTION

The coronavirus disease 2019 (COVID-19) has seriously affected different aspects of human life [9, 32, 33, 36, 42, 45, 47]. To mitigate its spread, decision-makers and public authorities have issued various policies [3, 12, 18, 19, 22, 30, 53]. Correspondingly, a natural question to ask is: *which policy is more effective to mitigate the spread of COVID-19 in a given context?* Various studies such as correlation analysis [25, 40] address this question from a statistical perspective. Such studies can only capture the statistical (but may not be causal) dependencies between the policies and the spread of COVID-19. Yet answering this question from a *causal* perspective is essential, as it can provide guidance to policymakers for addressing other pandemics or even further waves of the current one [10, 14, 39]. However, the gold-standard of causal effect estimation, i.e., a randomized controlled trial [7] is not applicable under pandemic circumstances due to ethical and practical issues [2]. Hence, the causal impact of different policies on the COVID-19 outbreak dynamics (e.g., the numbers of confirmed cases) is expected to be directly assessed with the observational data.

In this paper, we assess the causal impact of different COVID-19 policies on the outbreak dynamics with observational data. Specifically, we study this problem: given a specific time period, what is the causal effect of COVID-19 policies (*causes/treatments*) on the outbreak dynamics (*outcomes*) in each county (*instance*)? For example, in March 2020, how would the number of confirmed cases have been changed in Albemarle county of Virginia *because of* the state government having enacted social distancing policies? One key challenge of causal effect estimation from observational data is the existence of *unobserved confounders* (*confounders* are the factors that influence both the *treatment* and the *outcome*), which may bring *confounding bias* [49] into the estimation. For example, in a county where residents have a high vigilance towards COVID-19, the government may issue policies to enforce social distancing at an early stage of the pandemic, but residents in this county also tend to be more alert to COVID-19 and thus will have lower probability of infections, even in the absence of said policies. Here the vigilance of residents is a confounder, and if handled improperly, we may incorrectly take the statistical correlations between the presence of these policies and the outbreak dynamics as causal relations. Confounding bias can be eliminated by adjusting for all the confounders (i.e., controlling for confounders) [49]. However, most existing works [21, 44] of causal assessment of COVID-19

related policies from the observational data are based on the un-confoundedness assumption [49], i.e., all the confounders can be observed. Similarly, the widely used difference-in-differences (DID) methods [1] are based on the parallel trend assumption [15], i.e., there are no (unobserved) factors influencing both the treatment and the growth trend of the outcome over a time period. However, in real-world scenarios, these assumptions are difficult to be satisfied due to the existence of unobserved confounders. Many confounders are hard to be quantified, e.g., residents' vigilance in the previous example, and traditional culture, political ideologies, etc. Such unobserved confounders often yield biased estimation results [49, 51]. Furthermore, since the COVID-19 pandemic has lasted over a year, one new challenge is that the unobserved confounders may also be time-varying, e.g., residents' vigilance may be low at an early stage, but increases when the situation becomes more severe. Nevertheless, to the best of our knowledge, most of the current studies on COVID-19 related policies lack the capability to handle the above challenges [13].

To remedy these issues, we first adopt a weaker form of the unconfoundedness assumption [1], which enables us to capture the unobserved confounders from the proxy variables for them, i.e., the variables which have dependencies with the unobserved confounders. For example, certain confounders such as residents' vigilance in a county can be inferred from the popularity of web searches about COVID-19 related keywords on Google. Intuitively, the more vigilant people are, the more frequently they will search COVID-19 terms. Besides, residents' vigilance can also be inferred from the relational information among different counties, such as a county-to-county distance network. One potential reason is that neighboring counties tend to have more interactions and similar culture, thus their residents will have similar levels of vigilance. Historical information such as the adopted policies and the spread of COVID-19 at earlier time periods may also influence the current confounders such as residents' vigilance. With such proxy variables, the confounders can be captured and thus an unbiased causal effect estimation becomes possible.

Bearing this in mind, we integrate data from several different data sources, including 391 counties in the United States. We consider our selected counties representative as they cover different states with different political ideologies. Our data includes COVID-19 related policies as treatments, the number of confirmed cases or death cases as outcomes. Besides, it also includes multiple covariates for each county and relational information among different counties, which can serve as proxy variables for capturing unobserved confounders. To tackle our studied problem, we utilize the above observational data and propose a neural network based framework — **COVID-19 de**confounde**r** (*CIDER*), to learn the representations of (unobserved) confounders. With the learned representations, we estimate the causal effects of different policies on COVID-19 outbreak dynamics. In our proposed framework, a Recurrent Neural Network (RNN) [37] is utilized to capture temporal information, while a Graph Convolutional Network (GCN) [28] is used to handle the relational information among different counties. At each time period, we predict the outbreak dynamics and whether these policies are in effect in each county. In this way, we can effectively utilize the observational data and learn the representations of time-varying confounders for causal assessment of COVID-19 policies.
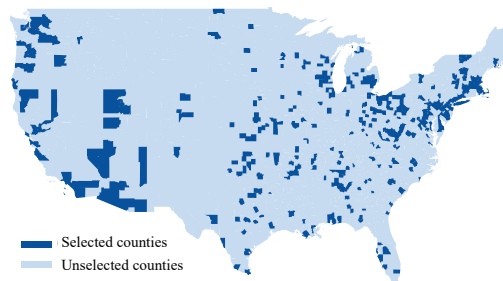

Figure 1: Geolocation of the selected counties in our corpus.

Our contributions can be summarized as: 1) We study the important problem of estimating the causal effect of COVID-19 policies on the outbreak dynamics for a given county and a specific time period. 2) We integrate data from several COVID-19 related sources. Then we develop a neural network based framework based on both relational and time-varying observational data to control for the unobserved confounders. 3) We assess the causal effect of different policies on the outbreak dynamics of COVID-19 with the developed framework. The assessment includes policies at different granularities, ranging from a category of policies with a certain goal to a specific policy type. 4) We conduct extensive experiments and find that our framework outperforms several alternatives and also yields important insights for current or future pandemics.

2 DATA AND ANALYSIS

In this section, we describe how we prepare data for assessing the causal impact of COVID-19 related policies on outbreak dynamics. Some preliminary data analyses are also presented to show the potential capability of capturing the (unobserved) confounders.

2.1 Observational Data

In general, two basic types of information are indispensable in the causal inference study, i.e., treatment and outcome. Specifically, for treatment, we collect COVID-19 related policies that have been enacted by different counties in the United States throughout 2020; for outcome, we use the numbers of confirmed cases and death cases of different counties throughout 2020. To control the influence of unobserved confounders, we also collect data regarding the covariates of different counties and their relations. In particular, two types of networks among counties are used in our study. In total, after filtering the counties without sufficient data, our data corpus includes 391 counties in the United States. The locations of these selected counties in our corpus are shown in Fig. 1. We then introduce how we collect and preprocess the data as follows.

Treatment — COVID-19 related policies. We collect the COVID-19 related policies that are in force in the USA throughout 2020 from the Department of Health & Human Services[1]. A total number of 60 policy types are included, along with descriptions and start/end dates. The collected policies include both state-level and county-level ones. For state-level policies, we assume that the policy applies to all counties in the state; for county-level policies, they are considered as only applying to the corresponding county. In order to better analyze the effect of these policies, we perform the

---

[1]https://catalog.data.gov/dataset/covid-19-state-and-county-policy-orders

following preprocessing: (1) *Policy filtering.* We remove the policy types that are adopted in less than 10% of the counties in our corpus, as we do not have sufficient observational data with respect to them. (2) *Policy categorization.* Based on the goal and some key element of each policy, we group them into three categories: *reduce contacts through social distancing* (henceforth referred to as social distancing), *minimize damage to the economy through reopening* (reopening), and *reduce airborne transmission through mask requirements* (mask requirements). For each category, Fig. 2 shows the top-10 policy types adopted by the largest number of counties, and the proportion of counties adopting them throughout 2020.

Outcome — numbers of confirmed and death cases. The daily numbers of confirmed and death cases are collected across these 391 counties from Johns Hopkins Coronavirus Resource Center[2] from January to December, 2020. In our experiments, we regard these numbers as outcomes of each county.

Covariates — popularity of keywords on Google. Unobserved confounders which causally affect the policies and outbreak dynamics are hard to capture. Hence, we use some proxy variables such as covariates of counties to infer these confounders. More specifically, we consider the web search of COVID-19 related keywords (e.g., coronavirus, mask, quarantine, etc.) by residents in these counties as covariates. We collect such web search data from Google Trends[3]. In this process, we first select a set of COVID-19 related keywords, and then obtain their *popularity score* based on the corresponding proportion in the total searches in each county from Google Trend. A higher popularity score implies that a larger proportion in this county has a high vigilance of COVID-19. In total, 19 different keywords are selected, and we obtain a 19-dimensional covariate vector for each county on each day from February to December 2020.

Networks — distance network and mobility flow network. Previous works [16, 17] have shown that network structure among instances can reflect some unobserved confounders. Therefore, in this work, two networks including the geographical distance network and mobility flow network among the selected 391 counties are collected as another kind of proxies for confounders. *1) Geographical distance network.* Geographical distance network among counties can be utilized to capture confounders. Intuitively, counties that are geographically closer are more likely to have similar confounders [16] such as residents' vigilance, because they tend to have similar cultural background and social climate. We construct a weighted network among the 391 counties based on the geographical distance from County Distance Database[4], where nodes represent counties, and edge weights are calculated from the corresponding distances between county pairs. Specifically, we select the county pairs with distance less than a threshold $\tau = 100$ kilometers, and set the weight as $1/d(i, j)$ for the edge between the i-th and j-th counties with distance $d(i, j)$. *2) Mobility flow network.* The mobility flow network among counties can also be adopted to capture confounders, as counties with large mobility flow are more likely to have similar confounders [16, 17] (such as residents' vigilance) as they have more communications. We construct an temporal mobility flow network with weighted edges among the 391 counties based on COVID19 USFlows [26], which tracks anonymous GPS

pings based on mobile applications. In this temporal network, the total daily volume of mobility flow is aggregated at different scales (e.g., census tract, county, and state) w.r.t. the timeline spanning from February to December in 2020. Each node denotes a county, and the weight of each edge is calculated from the mobility flow volume between the corresponding pair of counties. Specifically, we set the weight as $\frac{\log flow(i,j)}{\max_{i,j} \log flow(i,j)}$ for the edge between the i-th and j-th counties with mobility flow $flow(i, j)$.

2.2 Preliminary Data Analysis

To explore whether the covariates and networks have the potential to capture the (unobserved) confounders, we conduct preliminary data analysis to explore their dependencies with COVID-19 outbreak dynamics (i.e., the number of confirmed/death cases). Due to the space limit, we only show the analysis on the cumulative confirmed cases number of ten counties from Virginia (VA) and Massachusetts (MA) as examples. Similar observations can also be found in other counties, as well as the number of death cases.

Relations between covariates and outbreak dynamics. As proxy variables of unobserved confounders, covariates of counties could have dependencies with the COVID-19 outbreak dynamics (outcome). For example, counties with relatively higher similarity of covariates may also have higher similarity in the number of confirmed cases. If such dependencies exist, it implies that the covariates of counties could be potentially helpful in capturing the unobserved confounders. In this regard, we explore whether such dependency exists in our collected covariates of counties, i.e., popularity of COVID-19 related search keywords of counties on Google US. We first compute the bivariate correlation between the daily cumulative confirmed case number series of the chosen counties and all 391 counties in 2020. Then for each of the ten counties, we rank its bivariate correlation values with the 391 counties in an ascending order. The average bivariate correlation value over every 10 percentile of the ranking is reported in Fig. (3a). Due to space constraints, we explain the process here only for one of the 19 keywords we used, "social distance". Similar observations can also be drawn based on other keywords. For each county, we represent the daily popularity of "social distance" on Google US as a time series spanning from February to December 2020. The bivariate correlation between counties based on their daily popularity is then calculated. Following the same ranking order in each row of Fig. (3a), we report the average bivariate correlation value of their daily popularity over every 10 percentile in Fig. (3b) in exponential scale. The results implies that most county pairs with higher bivariate correlation w.r.t. outbreak dynamics tend to have higher keyword popularity similarity. This reveals that our collected covariates have the potential to help capture the confounders.

Relations between networks and outbreak dynamics. Networks could also have dependencies with COVID-19 outbreak dynamics, e.g., county pairs with relatively shorter distance or larger mobility flow volume may also have higher similarity in the number of confirmed cases. Similar to the relationship between covariates and outbreak dynamics, such dependencies imply the potential utility of networks in capturing the confounders. Consequently, here we explore whether such dependencies exist in networks among counties. Following the same ranking order in each row of Fig. (3a), we also report the corresponding value of their distance and
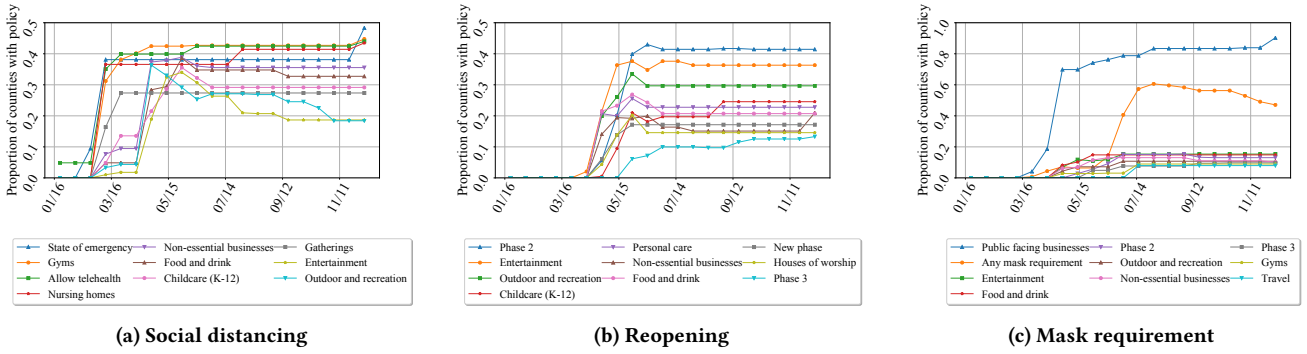
(a) Social distancing

(b) Reopening

(c) Mask requirement

Figure 2: Proportion of counties with policy types in each category over the course of 2020.

(a) Correlation between confirmed case numbers in different counties.

(b) Correlation between keyword popularity in different counties.

(c) Geographical distance between counties.

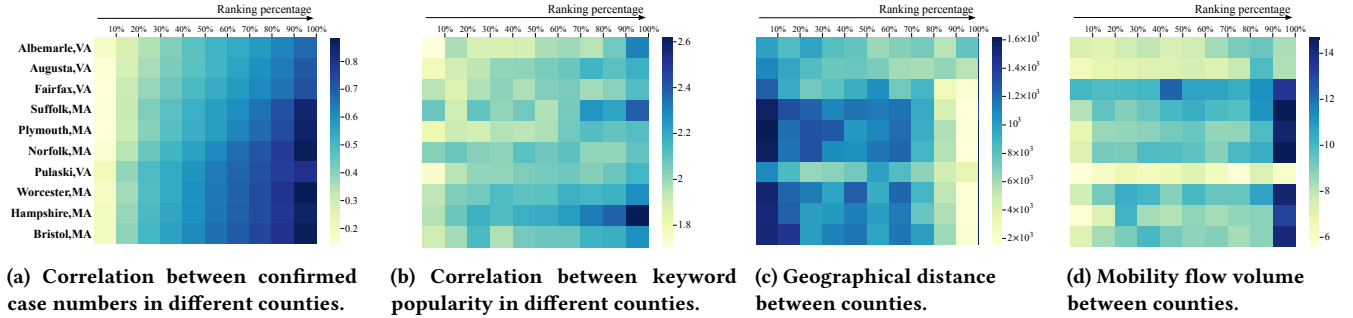**(d) Mobility flow volume between counties.**

Figure 3: Illustrations reflecting interactions between the selected counties and other counties: (a) bivariate correlation between the confirmed case number series in different counties, (b) bivariate correlation between keyword popularity in different counties, (c) geographic distance between counties, and (d) mobility flow volume between counties. The counties in each row of (a) are ranked by the bivariate correlation of the confirmed case number in an ascending order, and all the results are averaged over every 10 percentile of counties. Each row in (b), (c) and (d) follows the same order of the counties as in (a).

mobility flow (aggregated between Feb. and Dec. in 2020) averaged over every 10 percentile (in log scale) in Fig. (3c) and Fig. (3d), respectively. We have the following conclusions: 1) most county pairs with relatively lower bivariate correlation w.r.t. the outbreak dynamics are more likely to have larger distance than those with high correlation; 2) most county pairs with relatively higher bivariate correlation w.r.t. the outbreak dynamics tend to have larger human mobility flow volume between them. These imply the potential of the collected networks to capture the confounders.

3 TIME-VARYING CAUSAL ASSESSMENT

In this section, we formulate the causal assessment of COVID-19 related policy types as a causal effect estimation problem from observational data. To mitigate the confounding bias of such assessment, we develop a neural network based framework to capture the time-varying confounders.

3.1 Formulating Policy Assessment as A Causal Effect Estimation Problem

We consider the COVID-19 related policy types in n counties across T time periods. Different counties are described by the same set of covariates (i.e., features) over time, and we denote them by $X^t = \{x_i^t\}_{i=1}^n$, where x_i^t represents the covariates of the i-th county at time period t (e.g., in Albemarle county, VA, residents' search popularity of COVID-19 related keywords on Google throughout March, 2020). We represent the adjacency matrix of the network

(e.g., the distance network) at time period t as $A^t \in \mathbb{R}^{n \times n}$, where A_{ij}^t is the weight of edge $i \rightarrow j$ in A^t, and $A_{ij}^t = 0$ if there does not exist such edge. We assume that the edge weight can reflect the similarity between counties. For each policy type, we use $P^t = \{p_i^t\}_{i=1}^n$ to denote whether policies of this type are in effect in these n counties at time period t, where p_i^t is either 1 (treated) or 0 (not treated, i.e., controlled), corresponding to whether the policy type is in effect in the i-th county or not. At each time period, the treated counties form the *treated group*, while the controlled counties form the *control group*. Here, we denote a specific manifestation of outbreak dynamics (such as the number of confirmed cases) at time period t as $Y^t = \{y_i^t\}_{i=1}^n$, which are also referred to as *observed outcomes*. The history of covariates before time period t is denoted by $\overline{X}^t = (X^1, X^2, ..., X^{t-1})$, and the history of treatments $\overline{P}^t$ and network structures $\overline{A}^t$ are defined similarly. Then, we denote all the historical observational data before time period t as $\overline{\mathcal{H}}^t = \{\overline{X}^t, \overline{A}^t, \overline{P}^t, \overline{Y}^t\}$.

We frame the causal assessment of COVID-19 related policies as a causal effect estimation problem from time-varying observational data. Our goal is to investigate to what extent a *cause* (i.e., *treatment*, e.g., a policy in effect) would causally affect an *outcome* (e.g., the number of confirmed cases) for each instance (e.g., a county) at different time periods. To estimate the causal effect of a policy on the outbreak dynamics at time period t, we should compare the potential outcomes of the outbreak dynamics in each county if this
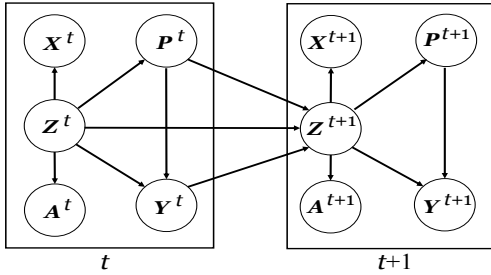
Figure 4: Causal graph of our studied problem.

policy had/had not been in effect during time period t. Generally, the *potential outcome* [43, 49] means the outcome that would be realized if the instance got treated/controlled, e.g., "In March, what would the number of confirmed cases be in Albemarle county, VA, if the mask requirement policy had/had not been in effect during that time period?". We denote the potential outcomes of all counties if the policy had/had not been in effect at time period t by $Y_1^t = \{y_{1,i}^t\}_{i=1}^n$ and $Y_0^t = \{y_{0,i}^t\}_{i=1}^n$. In our setting, the potential outcome $y_{p,i}^t$ is the outcome that would be realized if the i-th instance is under treatment p at time period t. Then the *individual treatment effect (ITE)* [49] for each instance at time period t is defined as the difference between the potential outcome if the instance gets treated and the potential outcome if it gets controlled at that time period. In our problem, the ITE [5] of each policy on the outbreak dynamics in each county at time period t is the difference between two potential outcomes: $\tau_i^t = \mathbb{E}[y_{1,i}^t - y_{0,i}^t | x_i^t, A^t, \overline{\mathcal{H}}^t]$. Then the average treatment effect (ATE) at time period t is computed as the average of ITEs over all counties: $\tau_{ATE}^t = \mathbb{E}_{i \in [n]}[\tau_i^t]$. Inspired by the previous work on causal effect estimation from time-evolving data [35], we design a causal graph for our studied problem as shown in Fig. 4, where each arrow means a causal relationship. We denote the time-varying (unobserved) confounders (e.g., residents' vigilance) by $Z^t = \{z_i^t\}_{i=1}^n$. In this paper, we use a weaker version of the unconfoundedness assumption [49] by assuming that there may exist unobserved confounders, and conditioning on all the confounders, the treatment assignment is independent with the potential outcomes, i.e., $y_{1,i}^t, y_{0,i}^t \perp\!\!\!\perp p_i^t | z_i^t$. Together with other commonly-used assumptions [24, 48, 49] in causal inference, we can identify and unbiasedly estimate the ITEs defined in our problem under the causal graph in Fig. 4 with a similar proof in [34].

3.2 CIDER: Estimating Causal Effects of COVID-19 Policies on Outbreak Dynamics

To better estimate the causal effects of COVID-19 related policies on the outbreak dynamics at each time period, we develop a neural network based framework CIDER with input as the time-varying observational data. Specifically, CIDER learns representations of time-varying confounders and conducts causal assessment based on these learned representations.

3.2.1 Learning Representations for Time-varying Confounders. Our framework captures the time-varying confounders from the evolving observational data at different time periods. As shown in Fig. 4,

we assume the confounders causally affect both the treatment assignment and the outcome at the current time period, while the treatment assignments, outcomes, and confounders at the previous time period also affect the current confounders. To capture the confounders from the time-varying observational data, we first use recurrent neural networks (RNNs) [37] to extract useful historical information from the data at previous time periods:

$$H^t = \text{RNN}(H^{t-1}, (Z^t \oplus X^t \oplus P^t \oplus Y^t)), \tag{1}$$

where H^t is the hidden unit in the RNNs which captures the historical information, including the previous treatment assignment $\overline{P}^t$ (i.e., whether a policy was in effect), outcomes $\overline{Y}^t$ (i.e., numbers of confirmed or death cases), network structure $\overline{A}^t$ (e.g., distance network) and observed features $\overline{X}^t$ (e.g., search popularity of COVID-19 related keywords). Here $\oplus$ denotes a concatenation operation. Besides, to better capture the unobserved confounders, we also take advantage of the network structure among counties using graph convolutional networks (GCNs) [28]. At each time period, we learn the representation of current confounders from the hidden state H^{t-1}, the current network structure A^t and covariates X^t:

$$Z^t = \widehat{A}^t \text{ReLU}(\widehat{A}^t (X^t \oplus H^{t-1})W_0)W_1, \tag{2}$$

where we stack two GCN layers to capture the non-linear dependencies between the unobserved confounders and the input data. W_0, W_1 are the parameters of the GCNs layers. $\widehat{A}^t$ is the normalized adjacency matrix computed from A^t beforehand with the renormalization trick [28]. Specifically, $\widehat{A}^t = (\widetilde{D}^t)^{-\frac{1}{2}} \widetilde{A}^t (\widetilde{D}^t)^{-\frac{1}{2}}$, where $\widetilde{A}^t = A^t + I_n$, $\widetilde{D}_{jj}^t = \sum_j \widetilde{A}_{jj}$. In this way, we learn the representation of the time-varying confounders. Notice that CIDER is not limited to specific proxy variables; it can utilize more proxy variables (if provided) to enhance the performance.

3.2.2 Balancing the Representations of Confounders. For certain COVID-19 related policies, the counties that had this policy (treated group) and those that did not (control group) often have different distributions of confounders. Such imbalanced distribution may bring additional biases for causal effect estimation [51]. Therefore, we employ representation balancing techniques at each time period by adding a distribution balancing constraint $\mathcal{L}_b$, which is the Wasserstein-1 distance of the representation distributions between the treated group and control group [51].

3.2.3 Prediction with Confounder Representations. To train a model that captures the representations of time-varying confounders, we utilize the observed outcome (i.e., outbreak dynamics) and treatment assignment (i.e., presence of policies) as supervision signals. Firstly, we use two multilayer perceptrons (MLPs) f_1 and f_0 to predict the potential outcomes:

$$\widehat{y}_{1,i}^t = f_1(z_i^t), \widehat{y}_{0,i}^t = f_0(z_i^t). \tag{3}$$

This way, each instance's *observed outcome* y_i^t (i.e., factual outcome, e.g., the observed number of confirmed cases) and *counterfactual outcome* (unobserved outcome with the contrary treatment, e.g., the potential number of confirmed cases if the policy had been treated differently from the fact) are estimated. The mean squared error (MSE) is used as the observed outcome prediction loss: $\mathcal{L}_y = \mathbb{E}_{t \in [T], i \in [n]}[(\widehat{y}_i^t - y_i^t)^2]$. Secondly, we also utilize the treatment

---
[5]Following [51], we define ITE in the form of the Conditional Average Treatment Effect (CATE) in this work.

Table 1: Examples of detailed policies about selected policy types in each category (including the states that enacted them). The three parts correspond to the categories of social distancing, reopening, and mask requirement, respectively.

Top policy types	Example policies
State of emergency	Limit in-person gatherings (VA)
Nursing homes	Limit visits to hospitals (OH)
Food and drink	Outdoor dining, and delivery only (RI)
Childcare (K-12)	Close public schools (VI)
Gatherings	Gatherings limited to 10 people (OH)
Phase 2	Reopen lodging establishments (NM)
Entertainment	Some businesses may reopen (ND)
Outdoor and recreation	Contact sport practices to reopen (OH)
Personal care	Health care may reopen (ND)
Food and drink	Reopen on-site dining (TN)
Any mask requirement	Face covering required in public (IN)
Public facing businesses	Facing businesses require masks (WI)
Food and drink	Restaurants require face masks (MP)
Phase 3	Outdoor venues require masks (ID)
Phase 2	Mandate masks in K-12 schools (UT)

assignment to train the representations of confounders. Specifically, we use a binary prediction module activated by a softmax layer. The output logit $\widehat{l}_i^t$ estimates the probability of getting treated, i.e., whether a policy is in effect in county i at time period t. The cross-entropy loss is adopted for the treatment assignment prediction:

$$\mathcal{L}_p = -\mathbb{E}_{t \in [T], i \in [n]}[(p_i^t \log(\widehat{l}_i^t) + (1 - p_i^t) \log(1 - \widehat{l}_i^t))]. \quad (4)$$

The overall loss function of our framework is:

$$\mathcal{L} = \mathcal{L}_y + \alpha \mathcal{L}_p + \beta \mathcal{L}_b + \lambda \|\Theta\|^2, \quad (5)$$

where α, β, λ are positive balancing hyperparameters, and Θ denotes the set of model parameters. After training the model, we estimate the ITE of the studied COVID-19 policy in each county i at time period t as the difference between the predicted potential outcomes $\widehat{\tau}_i^t = \widehat{y}_{1,i}^t - \widehat{y}_{0,i}^t$ and assess the ATE by averaging the estimated ITEs over all counties.

4 EXPERIMENTAL EVALUATIONS

In this section, we first show the estimated causal effects of different policies on the outbreak dynamics (number of confirmed/death cases) at different time periods during 2020. Based on the estimation, we assess the impact of these policies on both macro and micro levels. Then we evaluate our framework with respect to its ability of controlling for confounding. We also compare the causal effects estimated by different state-of-the-art methods for further discussion.

Setup. Counties are randomly assigned to 60%/20%/20% training/validation/test data. The results are averaged over 10 repeated executions. We set the length of each time period as 15 days, the learning rate as 0.003, epochs as 2000, the dimension size of the confounder representation $d_z = 50$, $\alpha = 1.0$, $\beta = 1e-4$, $\lambda = 0.01$. We use an Adam optimizer.

4.1 Causal Assessment of COVID-19 Policies

We estimate the causal effects of different policies on the outbreak dynamics at different time periods in 2020. Based on our estimation of their average treatment effects (ATEs), Table 1 shows the

information about the top-5 most impactful policy types in each policy category with a certain goal. Some policy types may contain policies in different categories. For example, in Table 1, for category Reopening, policy type "Phase 2" covers the reopening-related policies like "Reopen lodging establishments". For category Mask, "Phase 2" covers mask-related policies like "Mandate masks in schools". Fig. 5 summarizes our estimation of the ATEs of these policy types. Each column in Fig. 5 corresponds to a category: *social distancing*, *reopening*, and *mask requirement*. The first and second rows show the estimated ATEs of these policy types on the number of confirmed cases and death cases, respectively. We have the following observations from Fig. 5:

At the macro-level, the policy types regarding social distancing and mask requirement have negative causal effects on both the number of confirmed cases and death cases, while the policy types about reopening have positive causal effects. These negative values of causal effect indicate that the corresponding policy types causally help reduce the spread of COVID-19, while the policy types with positive causal effects may have a contrary effect because they increase the risk of infection. These observations appear intuitively plausible and are also consistent with existing literature regarding COVID-19 related policies [22, 41].

At the micro-level, we zoom into the most impactful policy types in each category. In the category of social distancing, the policy types "Gatherings" and "Food and drink" seem to have the strongest effects. From the detailed description of these policies, they powerfully prohibit the number of individuals in multiple activities, especially in high-risk places such as restaurants. In the category of reopening, the estimated effects of policy types "Personal care" and "Food and Drink" indicate that reopening public places such as personal care center and restaurants heavily increase the risk of COVID-19 infection. In the category of mask requirement, "Phrase 2" and "Any mask requirement" are most impactful as they mandate face masks usage in many public spaces.

Generally, above observations are consistent for different outcomes including the number of confirmed/death cases, as well as for different time periods. Besides, we observe that the policies can have stronger effects when the the pandemic becomes more severe, such as during the outbreak at the end of 2020. In conclusion, above observations reveal the importance of in-time policies to limit close contact (within about 6 feet) among people in different aspects, e.g., distance, frequency and certain body parts (e.g., face) during the spread of respiratory pandemics like COVID-19.

Furthermore, we zoom into the county-level, and assess the ITEs of different policy types on the outbreak dynamics in each county. In Fig 6, we randomly select 10 counties as examples and show the estimated ITEs of different policy types in each of them. To compare different counties, each result is calculated from the original ITEs (as defined in Eq. 3.1) averaged over all the time periods, and then is normalized by the number of confirmed cases in the corresponding county at the last time period. From Fig. 6, we observe that the ITEs of the three categories of policies in each county generally have similar patterns as their ATEs over all counties, i.e., the "social distancing" and "mask requirement" policies are beneficial for controlling the spread of COVID-19, while the "reopening" policies have increased the risk. Besides, the policies have a stronger impact in high-risk locations such as California, New York, and Florida.

(a) Social distancing, confirmed cases.   (b) Reopening, confirmed cases.   (c) Mask requirement, confirmed cases.

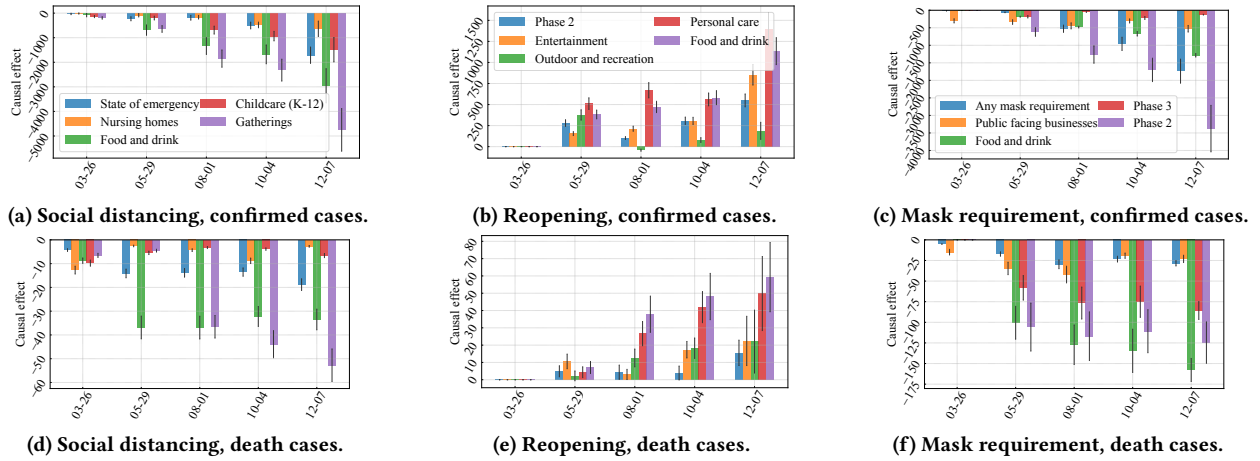(d) Social distancing, death cases.   (e) Reopening, death cases.   (f) Mask requirement, death cases.

Figure 5: Causal effect estimation of different policy types at different time periods over year 2020. The three columns correspond to the policy categories of social distancing, reopening, and mask requirements. The two rows correspond to the estimated causal effects on the number of confirmed cases and the number of death cases, respectively.

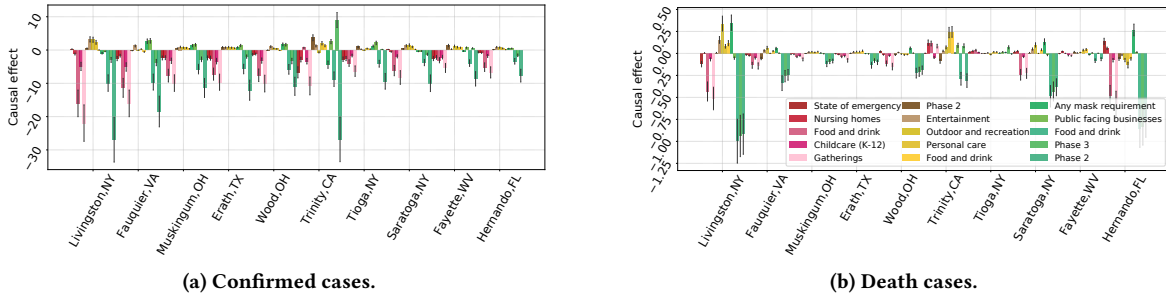(a) Confirmed cases.   (b) Death cases.

Figure 6: Causal effect estimation of different policy types on the outbreak dynamics in different counties. The red, yellow and green bars correspond to the policy categories of social distancing, reopening, and mask requirement, respectively.

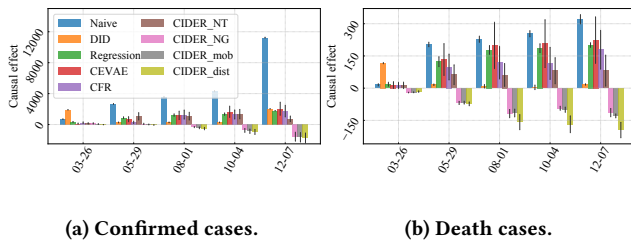## 4.2 Prediction with Learned Confounders

Due to the lack of counterfactual outcomes, it is hard to quantitatively evaluate the performance of causal effect estimation. Fortunately, the prediction performance of outbreak dynamics (outcome) and presence of policies (treatment) based on the learned confounder representations can serve as a good indicator to show the capability of the proposed framework in capturing the confounders and implicitly assessing the effectiveness of causal effect estimation. Here we compare the prediction performance and causal effect estimation results of our proposed framework with multiple baselines, including the state-of-the-art causal effect estimation methods, as well as some variants of our proposed framework for ablation study. The baselines include: (1) **Naive estimation of ATE** [49] — This method estimates the causal effect of each policy by simply taking the difference between the average values of the observed outcomes in the treated group and the control group. This method may suffer from confounding bias. (2) **Outcome regression** [49] — Outcome regression is a commonly used method in causal effect estimation. It takes the covariates as input to predict the potential outcomes under each treatment assignment. We implement it with linear regression. (3) **Difference-in-differences (DID)** — DID [1] estimates the causal effect by comparing the average change of the outcome during each time period in the treated group with the change in

the control group. It is based on the parallel trend assumption [15]. (4) **Causal effect variational autoencoder (CEVAE)** [34] — CEVAE is a deep latent-variable model, which learns representations of confounders as Gaussian distributions from original features, observed treatment assignment, and factual outcome. (5) **Counterfactual Regression (CFR)** [51] — CFR learns representation for the confounders based on the unconfoundedness assumption [49] and predicts the potential outcomes based on the learned representations. For all the above methods, we process each time period separately. (6) **No network (CIDER-NG)** — In this variant of our framework, we remove the GNN module to disable this variant from utilizing the network structure. (7) **No temporal (CIDER-NT)** — In this variant, we remove the RNN based memory unit to disable this variant from utilizing the temporal information. As the baselines Naive estimation of ATE, Outcome regression, CEVAE and CFR cannot handle the temporal information, we use them separately at each time period.

Table 2 compares the performance of different methods in predicting the treatment assignment (whether the policies are present) and outcomes (the numbers of confirmed/death cases). We denote our framework based on the distance network and mobility network with postfix -dist and -mob, respectively. As other methods do not model the treatment prediction, we compare our method and its

Table 2: Comparison of the performance of different methods for outcome and treatment predictions ("SD": social distancing, "RO": reopening, "MA": mask requirement).

	RMSE of outcome prediction		Accuracy of treatment prediction		
Methods	Confirmed	Death	SD	RO	MA
Regression	13145.34	395.83	-	-	-
CEVAE	12236.19	378.25	-	-	-
CFR	12577.42	316.46	-	-	-
CIDER-NT	11540.94	287.91	0.863	0.872	0.843
CIDER-NG	751.82	62.41	0.881	0.892	0.869
CIDER-dist	657.44	53.54	0.901	0.901	0.885
CIDER-mob	694.48	55.25	0.893	0.901	0.874

(a) Confirmed cases.

(b) Death cases.

Figure 7: Causal effect estimation of mask-related policy types computed by different methods.

variants in treatment prediction w.r.t. the policy types in different categories. The results show that our framework can achieve high performance in both outcome and treatment assignment predictions, indicating that our framework can capture the confounders well, as discussed in [51]. By comparing our framework and its two variants, we observe that the network and temporal information can both benefit the two kinds of predictions, suggesting that these information can help capture more (unobserved) confounders.

4.3 Causal Effect Estimation Benchmarking

To further investigate the capability of our framework in capturing the unobserved confounders, we conduct the following case study: we integrate the policies in the mask category as one treatment, and Fig. 7 shows the comparison of the causal effect estimated by different methods. We observe that the results estimated by almost all the baselines stay positive values over time, which conflicts with our common knowledge. This phenomenon may result from some unobserved confounders, e.g., people in the counties which are under more severe situations (e.g., higher infection rates in itself or neighboring counties) may prone to adopt the mask requirement earlier, while other counties may not be so alert to adopt these policies. Without effective ways of capturing the confounders (e.g., current situation), a method may incorrectly take the dependency between "wearing mask" and "high number of confirmed cases" as a causal relationship between them, and lead to such biased estimation. Our framework can capture the (unobserved) confounders, thus can provide estimation which is more consistent with common knowledge and epidemiological studies of COVID-19 [41].

5 RELATED WORK

Non-causal analysis of COVID-19 policies. Various studies have investigated the COVID-19 related policies [4, 6, 18, 22, 27, 44], including their relations with multiple social and economic factors such as household incomes [5, 46], economics [21, 50], and vaccination [38, 40, 52]. For example, some studies [25] used correlation analysis or simulation methods to show the effectiveness of different COVID-19 policies in controlling this pandemic. Another work proposed a conceptual framework [21] to analyze how non-pharmaceutical interventions such as social distancing will impact economics. However, these works can only reveal the statistical dependencies between the policies and other factors, rather than identifying the causal effects among them.

Causal analysis of COVID-19 policies. Different from the statistical perspective, various studies [22, 31] estimated the causal effect of different policies on COVID-19 dynamics. Among them, some works [11] are based on structural causal models (SCMs) [23]. These works need domain knowledge or causal discovery methods [20] to obtain a complete causal model that describe the causal relationships among different variables. However, it is often difficult to obtain a complete causal graph. Of the methods which do not assume a complete causal graph, DID methods [29] estimate the causal effect of a COVID-19 related policy by comparing the average change of outcome over time in the treated group with the control group. Other studies [8, 41] that are based on synthetic control methods can account for the effects of evolving confounders over time. Despite the contributions of these works on COVID-19 policy assessment, most of them are based on strong assumptions regarding the existence of unobserved confounders.

6 CONCLUSION

In this paper, we study the problem of assessing the causal impact of various COVID-19 related policies on the outbreak dynamics in different U.S. counties at different time periods. The main challenge here is the existence of unobserved and time-varying confounders. To address this problem, we integrate data from multiple COVID-19 related data sources containing different kinds of information which can serve as proxies for confounders. We develop a framework which learns the representations of the confounders by utilizing relational and time-varying observational data. Based on the learned representations, we assess the causal impact of these policies at different granularities. We also compare the prediction performance of outbreak dynamics and the presence of policies, as well as the causal effect estimation results of our framework with other baselines. The results implicitly validate the capability of our framework in controlling confounders for causal assessment of COVID-19 related policies. Future causal inference studies for COVID-19 analysis can address more problems in applications, e.g., the reported confirmed/death cases may contain noises or biases; some policies are cumulative and the treatment assignments should be considered as continuous variables rather than binary ones.

REFERENCES

[1] Alberto Abadie. 2005. Semiparametric difference-in-differences estimators. *The Review of Economic Studies* (2005).

[2] Clement Adebamowo, Oumou Bah-Sow, Fred Binka, Roberto Bruzzone, Arthur Caplan, Jean-François Delfraissy, David Heymann, Peter Horby, Pontiano Kaleebu, Jean-Jacques Muyembe Tamfum, et al. 2014. Randomised controlled trials for Ebola: practical and ethical issues. *The Lancet* (2014).

[3] Kavita Shah Arora, Jaclyn T Mauch, and Kelly Smith Gibson. 2020. Labor and delivery visitor policies during the COVID-19 pandemic: balancing risks and benefits. *Jama* (2020).

[4] Ben Balmford, James D Annan, Julia C Hargreaves, Marina Altoè, and Ian J Bateman. 2020. Cross-country comparisons of COVID-19: Policy, politics and the price of life. *Environmental and Resource Economics* (2020).

[5] Mike Brewer and Laura Gardiner. 2020. The initial impact of COVID-19 and policy responses on household incomes. *Oxford Review of Economic Policy* (2020).

[6] Giliberto Capano, Michael Howlett, Darryl SL Jarvis, M Ramesh, and Nihit Goyal. 2020. Mobilizing policy (in) capacity to fight COVID-19: Understanding variations in state responses. *Policy and Society* (2020).

[7] Thomas C Chalmers, Harry Smith Jr, Bradley Blackburn, Bernard Silverman, Biruta Schroeder, Dinah Reitman, and Alexander Ambroz. 1981. A method for assessing the quality of a randomized control trial. *Controlled clinical trials* (1981).

[8] Matthew A Cole, Robert JR Elliott, and Bowen Liu. 2020. The impact of the Wuhan Covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. *Environmental and Resource Economics* (2020).

[9] John Daniel. 2020. Education and the COVID-19 pandemic. *Prospects* (2020).

[10] Zeynep Ertem, Ozgur M Araz, and Mayteé Cruz-Aponte. 2021. A decision analytic approach for social distancing policies during early stages of COVID-19 pandemic. *Decision Support Systems* (2021).

[11] Navid Feroze. 2020. Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models. *Chaos, Solitons & Fractals* (2020).

[12] Andrea Galimberti, Hellas Cena, Luca Campone, Emanuele Ferri, Mario Dell'Agli, Enrico Sangiovanni, Michael Belingheri, Michele Augusto Riva, Maurizio Casiraghi, and Massimo Labra. 2020. Rethinking urban and food policies to improve citizens safety after COVID-19 pandemic. *Frontiers in Nutrition* (2020).

[13] Oguzhan Gencoglu and Mathias Gruber. 2020. Causal modeling of Twitter activity during COVID-19. *Computation* (2020).

[14] Bert George, Bram Verschuere, Ellen Wayenberg, and Bishoy Louis Zaki. 2020. A guide to benchmarking COVID-19 performance data. *Public Administration Review* (2020).

[15] Andrew Goodman-Bacon and Jan Marcus. 2020. Using difference-in-differences to identify causal effects of COVID-19 policies. (2020).

[16] Ruocheng Guo, Jundong Li, and Huan Liu. 2020. Learning individual causal effects from networked observational data. In *ACM International Conference on Web Search and Data Mining*.

[17] Ruocheng Guo, Yichuan Li, Jundong Li, K Selçuk Candan, Adrienne Raglin, and Huan Liu. 2020. IGNITE: A Minimax Game Toward Learning Individual Treatment Effects from Networked Observational Data. In *Proceedings of the 2020 International Joint Conference on Artificial Intelligence*.

[18] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al. 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* (2021).

[19] Daniel T Halperin, Norman Hearst, Stephen Hodgins, Robert C Bailey, Jeffrey D Klausner, Helen Jackson, Richard G Wamai, Joseph A Ladapo, Mead Over, Stefan Baral, et al. 2021. Revisiting COVID-19 policies: 10 evidence-based recommendations for where to go from here. *BMC public health* (2021).

[20] David Heckerman, Christopher Meek, and Gregory Cooper. 1999. A Bayesian approach to causal discovery. *Computation, causation, and discovery* (1999).

[21] Constantino Hevia and Andy Neumeyer. 2020. A conceptual framework for analyzing the economic impact of COVID-19 and its policy implications. *UNDP LAC COVID-19 Policy Documents Series* (2020).

[22] Solomon Hsiang, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, et al. 2020. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* (2020).

[23] Yimin Huang and Marco Valtorta. 2012. Pearl's calculus of intervention is complete. *arXiv preprint arXiv:1206.6831* (2012).

[24] Kosuke Imai and David A Van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *J. Amer. Statist. Assoc.* (2004).

[25] Jiwei Jia, Jian Ding, Siyu Liu, Guidong Liao, Jingzhi Li, Ben Duan, Guoqing Wang, and Ran Zhang. 2020. Modeling the control of COVID-19: impact of policy interventions and meteorological factors. *arXiv preprint arXiv:2003.02985* (2020).

[26] Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, and Jake Kruse. 2020. Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic. *Scientific Data* (2020).

[27] İlker Kayı and Sibel Sakarya. 2020. Policy analysis of suppression and mitigation strategies in the management of an outbreak through the example of COVID-19 pandemic. *Infect Dis Clin Microbiol* (2020).

[28] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[29] Kourtney Koebel and Dionne Pohler. 2020. *Labor markets in crisis: The causal impact of Canada's COVD19 economic shutdown on hours worked for workers across the earnings distribution.* Technical Report. Working Paper Series.

[30] Edward Kong and Daniel Prinz. 2020. Disentangling policy effects using proxy data: Which shutdown policies affected unemployment during the COVID-19 pandemic? *Journal of Public Economics* (2020).

[31] Werner Kristjanpoller, Kevin Michell, and Marcel C Minutolo. 2021. A causal framework to determine the effectiveness of dynamic quarantine policy to mitigate COVID-19. (2021), 107241.

[32] Sijia Li, Yilin Wang, Jia Xue, Nan Zhao, and Tingshao Zhu. 2020. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. *International journal of environmental research and public health* (2020).

[33] Shabir Ahmad Lone and Aijaz Ahmad. 2020. COVID-19 pandemic–an African perspective. *Emerging microbes & infections* (2020).

[34] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*.

[35] Jing Ma, Ruocheng Guo, Chen Chen, Aidong Zhang, and Jundong Li. 2021. Deconfounding with Networked Observational Data in a Dynamic Environment. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*.

[36] Mirko Manchia, Anouk W Gathier, Hale Yapici-Eser, Mathias V Schmidt, Dominique de Quervain, Therese van Amelsvoort, Jonathan I Bisson, John F Cryan, Oliver D Howes, Luisa Pinto, et al. 2022. The impact of the prolonged COVID-19 pandemic on stress resilience and mental health: A critical review across waves. *European Neuropsychopharmacology* (2022).

[37] Larry R Medsker and LC Jain. 2001. Recurrent neural networks. *Design and Applications* (2001).

[38] Jitendra Meena, Arushi Yadav, and Jogender Kumar. 2020. BCG vaccination policy and protection against COVID-19. *The Indian Journal of Pediatrics* (2020).

[39] Andrea Riccardo Migone. 2020. The influence of national policy characteristics on COVID-19 containment policies: A comparative analysis. *Policy Design and Practice* (2020).

[40] Aaron Miller, Mac Josh Reandelar, Kimberly Fasciglione, Violeta Roumenova, Yan Li, and Gonzalo H Otazu. 2020. Correlation between universal BCG vaccination policy and reduced mortality for COVID-19. *MedRxiv* (2020).

[41] Timo Mitze, Reinhold Kosfeld, Johannes Rode, et al. 2020. Face masks considerably reduce Covid-19 cases in Germany-A synthetic control method approach. (2020).

[42] M Mofijur, IM Rizwanul Fattah, Md Asraful Alam, ABM Saiful Islam, Hwai Chyuan Ong, SM Ashrafur Rahman, Gholamhassan Najafi, Shams Forruque Ahmed, Md Alhaz Uddin, and Teuku Meurah Indra Mahlia. 2021. Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic. *Sustainable production and consumption* (2021).

[43] Jersey Neyman. 1923. Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych* (1923).

[44] Peterson Ozili. 2020. COVID-19 in Africa: socio-economic impact, policy response and opportunities. *International Journal of Sociology and Social Policy* (2020).

[45] Betty Pfefferbaum and Carol S North. 2020. Mental health and the Covid-19 pandemic. *New England Journal of Medicine* (2020).

[46] Suphanit Piyapromdee and Peter Spittal. 2020. The income and consumption effects of covid-19 and the role of public policy. *Fiscal Studies* (2020).

[47] John L Romano. 2020. Politics of Prevention: Reflections From the COVID-19 Pandemic. *Journal of Prevention and Health Promotion* (2020).

[48] Donald B Rubin. 1980. Randomization analysis of experimental data: The Fisher randomization test comment. *J. Amer. Statist. Assoc.* (1980).

[49] Donald B Rubin. 2005. Bayesian inference for causal effects. *Handbook of Statistics* (2005).

[50] Falk Schwendicke, Joachim Krois, and Jesus Gomez. 2020. Impact of SARS-CoV2 (Covid-19) on dental practices: Economic analysis. *Journal of Dentistry* (2020).

[51] Uri Shalit, Fredrik D Johansson, and David Sontag. 2017. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*.

[52] Abhibhav Sharma, Saurabh Kumar Sharma, Yufang Shi, Enrico Bucci, Ernesto Carafoli, Gerry Melino, Arnab Bhattacherjee, and Gobardhan Das. 2020. BCG vaccination policy and preventive chloroquine usage: do they have an impact on COVID-19 pandemic? *Cell death & disease* (2020).

[53] Fernando A Wilson and Jim P Stimpson. 2020. US policies increase vulnerability of immigrant communities to the COVID-19 pandemic. *Annals of global health* (2020).