Causal Analysis of Health Interventions and Environments for Influencing the Spread of COVID-19 in the United States of America

Zhouxuan Li[1], Tao Xu[1], Kai Zhang[1], Hong-Wen Deng[2], Eric Boerwinkle[1] and Momiao Xiong[1]*

[1]School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States, [2]Tulane Center of Biomedical Informatics and Genomics, Deming Department of Medicine, School of Medicine, Tulane University, New Orleans, LA, United States

Given the lack of potential vaccines and effective medications, non-pharmaceutical interventions are the major option to curtail the spread of COVID-19. An accurate estimate of the potential impact of different non-pharmaceutical measures on containing, and identify risk factors influencing the spread of COVID-19 is crucial for planning the most effective interventions to curb the spread of COVID-19 and to reduce the deaths. Additive model-based bivariate causal discovery for scalar factors and multivariate Granger causality tests for time series factors are applied to the surveillance data of lab-confirmed Covid-19 cases in the US, University of Maryland Data (UMD) data, and Google mobility data from March 5, 2020 to August 25, 2020 in order to evaluate the contributions of social-biological factors, economics, the Google mobility indexes, and the rate of the virus test to the number of the new cases and number of deaths from COVID-19. We found that active cases/1,000 people, workplaces, tests done/1,000 people, imported COVID-19 cases, unemployment rate and unemployment claims/1,000 people, mobility trends for places of residence (residential), retail and test capacity were the popular significant risk factor for the new cases of COVID-19, and that active cases/1,000 people, workplaces, residential, unemployment rate, imported COVID cases, unemployment claims/1,000 people, transit stations, mobility trends (transit), tests done/1,000 people, grocery, testing capacity, retail, percentage of change in consumption, percentage of working from home were the popular significant risk factor for the deaths of COVID-19. We observed that no metrics showed significant evidence in mitigating the COVID-19 epidemic in FL and only a few metrics showed evidence in reducing the number of new cases of COVID-19 in AZ, NY and TX. Our results showed that the majority of non-pharmaceutical interventions had a large effect on slowing the transmission and reducing deaths, and that health interventions were still needed to contain COVID-19.

Keywords: COVID-19, causal inference, time series, control of the spread, transmission dynamics, public health interventions

INTRODUCTION

As of August 25, 2020, the number of cumulative cases of COVID-19 in the US exceeded 5,727,107 and included 170,305, deaths (John Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/MAP.HTML), thus causing a devastating public health and economic crisis. Since the number of new cases in the US remains high (36,339 in the US on August 25, 2020) ((John Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/MAP.HTML), curbing the spread of COVID-19 is urgently needed [1]. There is increasing recognition that many geographic, economic and environmental factors contribute to the outbreak of COVID-19. In the absence of vaccines and specifically effective medications, non-pharmaceutical public health interventions and personal hygiene practices are the only options to slow the spread of COVID-19 [2, 3]. The effects of the different factors and intervention measures on the spread of COVID-19 vary. Identifying key factors that most contribute to the rapid spread of COVID-19, and accurately estimating the potential impact of different non-pharmaceutical measures for containing COVID-19 are crucial for planning the most effective interventions to curb the spread of COVID-19 [4].

The widely used statistical methods for COVID-19 epidemiological factor analysis and evaluation of intervention measures include correlation analysis [2, 4–6], regression [7, 8], logistic regression [9] and a transmission dynamic model coupled with a linear model [10]. The most examined scalar factors consist of underlying health conditions such as high blood pressure, diabetes, stroke, cardiac or kidney diseases, and aging individuals [2, 11, 12], atmospheric temperature [5], age, gender, ethnicity, and population density [2, 13], airflow [2], and socioeconomics such as median income [9, 14].

The most explored non-pharmaceutical public health interventions and digital technologies for curbing the spread of COVID-19 include social distancing, case isolation and quarantine as well as closuring borders, schools travel restrictions, use of face-masks, and testing [15–18] and population surveillance, case identification, contact tracing, mobility data collection, and communication technology, which utilize billions of mobile phones and large online datasets to provide information for the evaluation of intervention strategies and to strengthen the curb of the spread of COVID-19 [17–19].

Although association analysis is of great importance for curbing the spread of COVID-19, association measures dependence between two variables or two sets of variables in the data, and use the dependence for prediction and evaluation of the effects of environmental, social-economic factors and public health interventions on the spread of COVID-19 [20, 21]. It is well recognized that association analysis is not a direct method to discover the causal mechanism of complex diseases. Association analysis may detect superficial patterns between intervention measures and transmission variables of COVID-19. Its signals provide limited information on the causal mechanism of the transmission dynamics of COVID-19 [22]. Association analysis has been a major paradigm for statistical

evaluation of the effects of influencing factors and health interventions on the spread of COVID-19 [23]. Understanding the transmission mechanism of COVID-19 based on association analysis remains elusive. The question to uncover the transmission mechanisms of COVID-19 is causal in nature.

Distinguishing causation from association is an age-old problem. Methods for causation analysis that is one of the most challenging problems in science and technology need to be developed as an alternative to association analysis [24]. A number of researchers have performed causal analysis of COVID-19 to evaluate the causal effects of mobility, awareness, and temperature [22], social distancing [21], mobility [25], herd immunity [26], and mask use [27]. However, most causal analysis of COVID-19 have treated time series data as pseudo-cross-sectional data. In some cases, causal analysis of COVID-19 treated the data as time series; time series was assumed stationary. In practice, the number of new cases and the number of deaths from COVID-19 were nonstationary time series in most cases [28]. The environmental, social-economic and geographic factors, and intervention measures include two types of data: scalar variables and time series (stationary or nonstationary) variables.

The purpose of this paper is to develop a general framework for the causal analysis of COVID-19 in the US. The number of new cases and deaths from COVID-19 are taken as response variables. The factors and intervention measures are taken as potential causal variables. If the factor and intervention variables are scalar variables, the additive noise models (ANMs) [29] are used to test for causation between the response variable and potential causal variable where the number of new cases or deaths should be averaged over time. Most intervention measures are time series data. An essential difference between time series and cross-sectional data is that the time series data have temporal order, but cross-sectional data do not have any order. As a consequence, the causal inference methods for cross sectional data cannot be directly applied to time series data. Basic tools in statistical analysis are the raw of large numbers and the central limit theorem. Applications of these tools usually assume that all moment functions are constant. When the moment functions of the time series vary over time, the raw of large numbers and the central limit theorem cannot be applied. In order to use basic probabilistic and statistical theories, the nonstationary time series must be transformed to stationary time series [30].

A widely used concept of causality for time series data is Granger causality [31, 32]. Underlying the Granger causality is the following two principles:

(1) Effect does not precede the cause in time;
(2) The effect series contains unique causal series information which is not present elsewhere.

The multivariate linear Granger causality test will be used to test causality between the number of new cases and deaths from COVID-19 and environmental, economic and intervention time

series variables [33]. The proposed ANMs and multivariate linear Granger causality analysis methods are applied to the surveillance data of lab-confirmed COVID-19 cases in the US, UMD data, and Google mobility data from March 5, 2020 to August 25, 2020 in order to evaluate the contributions of social-biological factors, economics, the Google mobility indexes, and the rate of virus testing to the number of the new cases and number of deaths from COVID-19. Data and software for implementing the algorithms for causal analysis can be downloaded from our website https://sph.uth.edu/research/centers/hgc/software/xiong/.

MATERIAL AND METHODS

Nonlinear Additive Noise Models for Bivariate Causal Discovery

The ANMs are used for identifying causal effect of a factor or an intervention measure on the number of new cases or an intervention measure [29, 34]. Assume no confounding, no selection bias, and no feedback. Let Y be the average number of new cases or new deaths from COVID-19 and X be a scalar factor or an intervention measure such as gender, population density, ethnic group, among others. Consider a bivariate additive noise model $X \to Y$ where Y is a nonlinear function of X and independent additive noise E_Y:

$$Y = f_Y(X) + E_Y, \quad X \coprod E_Y, \qquad (1)$$

where X and E_Y are independent. Then, the density $P_{X,Y}$ is said to be induced by the additive noise models (ANM) from X to Y [35]. In some cases, we may have the following alternative direction of the ANMs: $Y \to X$:

$$X = f_X(Y) + E_X, \quad Y \coprod E_X, \qquad (2)$$

where Y and E_X are independent. If the density $P_{X,Y}$ is induced by the ANM $X \to Y$, but not by the ANM $Y \to X$, then the ANM $X \to Y$ is identifiable.

Assume that $n + m$ state data were sampled. Divide the dataset into a training data set by specifying $D_1 = \{Y_n, X_n\}$, $Y_n = [y_1, \ldots, y_n]^T$, $X_n = [x_1, \ldots, x_n]^T$ for fitting the model and a test data set $D_2 = \{\tilde{Y}_m, \tilde{X}_m\}$, $\tilde{Y}_m = [\tilde{y}_1, \ldots, \tilde{y}_m]^T$, $\tilde{X}_m = [\tilde{x}_1, \ldots, \tilde{x}_m]^T$ for testing the independence, where n is not necessarily equal to m.

Procedures for using the ANM to assess causal relationships between two variables are summarized below [34].

Step 1. Regress Y on X using the training dataset D_1 and non-parametric regression methods:

$$Y = \hat{f}_Y(X) + E_Y \qquad (3)$$

Step 2. Calculate the residual $\hat{E}_Y = Y - \hat{f}_Y(X)$ using the test dataset D_2 and test whether the residual $\hat{E}_Y$ is independent of causal X to assess the ANM $X \to Y$.

Step 3. Repeat the procedure to assess the ANM $\to X$.

Step 4. If the ANM in one direction is accepted and the ANM in the other is rejected, then the former is inferred as the causal direction.

There are many non-parametric methods that can be used to regress Y on X or regress X on Y. For example, we can use neural networks [36], smoothing spline regression methods [37], B-spline [38] and local polynomial regression [39]. In this paper, the smoothing spline regression method was used to fit the regression models.

Covariance can be used to measure association but cannot be used to test independence between two variables with a non-Gaussian distribution (https://en.wikipedia.org/wiki/Correlation_and_dependence). A covariance operator that is a generalization of the finite dimensional covariance matrix to infinite dimensional feature space can be used to test for independence between two variables with arbitrary distributions. Specifically, we will use the Hilbert-Schmidt norm of the cross-covariance operator or its approximation, the Hilbert-Schmidt independence criterion (HSIC) to measure the degree of dependence between the residuals and potential causal variable and test for their independence [35, 40].

The covariance operator can be defined as

$$Cov\left(f(X), g(Y)\right) == \langle f, C(X, Y)g \rangle_{\mathcal{F}},$$

where f, g are any nonlinear functions and $C(X, Y)$ is the covariance operator and $\langle . \rangle_{\mathcal{F}}$ is an inner product in the $\mathcal{F}$ Hilbert space. The covariance operator is positive symmetric, which implies linearity and continuity. In addition, the covariance operator maps spaces in their dual spaces [41]. The Hilbert-Schmidt norm of the covariance operator can be used as criterion for assessing independence between two random variables and is called the Hilbert-Schmidt independence criterion (HSIC). The Hilbert-Schmidt norm of the centered covariance operator is defined as

$$HSIC^2(X, Y) = \|C_{XY}\|_{HS}^2,$$

where $\|.\|_{HS}$ is the Hilbert-Schmidt norm.

We know that (Wang et al., 2018). $HSIC^2(X, Y) = 0$ if and only if X and Y are independent. $HSIC^2(X, Y)$ can be approximated by.

$$HSIC^2(X, Y) = \frac{1}{n^2}\text{tr}(KHLH),$$

where n is a sample size, K and $L =$ are $n \times n$ dimensional kernel matrices and $H = I - \frac{1_n 1_n^T}{n}$. We used the Gaussian kernel: $k(x, y) = e^{-\frac{x-y^2}{2\sigma^2}}$, $\sigma > 0$ and polynomial kernel. The order-d polynomial kernel is defined as $k(x, y) = (x^T y + c)^d$. The fourth and sixth order polynomial kernels were used in this analysis. To test independence between the potential cause X and residual E_Y, we calculated $HSIC^2(X, E_Y)$ as follows.

$$HSIC^2(X, E_Y) = \frac{1}{n^2}\text{tr}(K_X H K_{E_Y}),$$

$$K_{E_Y} = \begin{bmatrix} k_{E_Y}(\varepsilon_1, \varepsilon_1) & \cdots & k_{E_Y}(\varepsilon_1, \varepsilon_n) \\ \vdots & \vdots & \vdots \\ k_{E_Y}(\varepsilon_n, \varepsilon_1) & \cdots & k_{E_Y}(\varepsilon_n, \varepsilon_n) \end{bmatrix}$$

$$\text{and } K_X = \begin{bmatrix} k_X(x_1, x_1) & \cdots & k_X(x_1, x_n) \\ \vdots & \vdots & \vdots \\ k_X(x_n, x_1) & \cdots & k_X(x_n, x_n) \end{bmatrix},$$

$$\varepsilon_i = \widehat{E}_{Y(i)} = Y_i - \widehat{f}_Y(X_i), \quad i = 1, \ldots, n,$$

In summary, the general procedure for testing independence between the average number of new cases or new deaths and the scalar factor or intervention measure is given as follows [34, 35]:

Step 1: Divide a data set into a training data set $D_n = \{Y_n, X_n\}$ for fitting the model and a test data set $D_m = \{\tilde{Y}_n, \tilde{X}_n\}$ for testing the independence.

Step 2: Use the training data set and any non-parametric regression methods

(a) Regress Y on X: $Y = f_Y(X) + E_Y$,
(b) Regress X on Y: $X = f_X(y) + E_X$.

Step 3: Use the test data set and any non-parametric regression methods that fits the test data set $D_m = \{\tilde{Y}_n, \tilde{X}_n\}$ to predict residuals:

(a) $\widehat{E}_{Y_X} = \tilde{Y} - \tilde{f}_Y(\tilde{X})$,
(b) $\widehat{E}_{X_Y} = \tilde{X} - \widehat{f}_X(\tilde{X})$.

Step 4: Calculate the dependence measures $HSIC^2(E_Y, X)$ and $HSIC^2(E_X, Y)$.
Step 5: Infer causal direction:

$$X \to Y \text{ if } HSIC^2(E_Y, X) < HSIC^2(E_X, Y); \quad (4)$$

$$Y \to X \text{ if } HSIC^2(E_Y, X) > HSIC^2(E_X, Y); \quad (5)$$

If $HSIC^2(E_Y, X) = HSIC^2(E_X, Y)$, then causal direction is undecided.

We do not have closed analytical forms for the asymptotic null distribution of the HSIC and hence it is difficult to calculate the p-values of the independence tests. To solve this problem, the permutation/bootstrap approach can be used to calculate the p-values of the causal test statistics. The null hypothesis is H_0 : no causations $X \to Y$ and $Y \to X$ (Both X and E_Y are dependent, and Y and E_X are dependent).

Calculate the test statistic

$$T_C = \left| HSIC^2(E_Y, X) - HSIC^2(E_X, Y) \right|. \quad (6)$$

Assume that the total number of permutations is n_p. For each permutation, we fix $x_i, i = 1, \ldots, m$ and randomly permute $y_i, i = 1, \ldots, m$. Then, fit the ANMs and calculate the residuals $E_X^i, E_Y^i, i = 1, \ldots, n$ and test statistic T_c. Repeat above procedures n_p times. The p-values are defined as the proportions of the statistic T_C (computed on the permuted data) greater than or equal to $\widehat{T}_C$ (computed on the original test data D_m).

Each state was a sample. Since the sample sizes were small (only 50), the p-value for declaring significance was 0.05 without Bonferroni correction for multiple comparison.

Multivariate Linear Granger Causality Test

Before performing multivariate linear Grander causality test, we first need to transform nonstationary time series to stationary time series.

Consider an m-variable VAR with p lags:

$$Y_t = \mu + \sum_{i=1}^{p} A_i Y_{t-i} + \varepsilon_t, \quad (7)$$

where Y_t is a m dimensional vector, μ is a mean, the A_i $(i = 1, \ldots, p)$ are $m \times m$ coefficient matrices and m dimensional residual vector ε_t is assumed to have mean zero $(E[\varepsilon_t] = 0)$, with no autocorrelation $(E[\varepsilon_t \varepsilon_{t-s}^T] = 0)$, but can be correlated across equations $(E[\varepsilon_t \varepsilon_t^T] = \Sigma)$.

Vector error correction model (VECM) consists of first differences of cointegrated $I(1)$ variables, their lags, and error correction terms:

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \varepsilon_t, \quad (8)$$

where matrixes Π and Φ_i $(i = 1, \ldots, p-1)$ are functions of matrices A_i $(i = 1, \ldots, p)$.

When two non-stationary variables are cointegrated, the VAR model should be augmented with an error correction term for testing the Granger causality (Engle and Granger, 1987).

The VECM can be reduced to

$$\Delta Y_t = A_0 + A(L) \Delta Y_{t-1} + \alpha(ECM_{t-1}) + \varepsilon_t, \quad (9)$$

where

$$\Delta Y_t = \begin{bmatrix} \Delta Y_{1t} \\ \vdots \\ \Delta Y_{mt} \end{bmatrix}, A_0 = \begin{bmatrix} A_{10} \\ \vdots \\ A_{m0} \end{bmatrix}, A(L) = \begin{bmatrix} A_{11}(L) & \cdots & A_{1m}(L) \\ \vdots & \vdots & \vdots \\ A_{m1}(L) & \cdots & A_{mm(L)} \end{bmatrix},$$

$$\alpha = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1k} \\ \vdots & \vdots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mk} \end{bmatrix},$$

$$ECM_{t-1} = \beta^T Y_{t-1} + Ct, \ Y_{t-1} = \begin{bmatrix} Y_{1,t-1} \\ \vdots \\ Y_{m,t-1} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \vdots & \vdots \\ \beta_{m1} & \cdots & \beta_{mk} \end{bmatrix}, C = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}.$$

Consider two non-stationary time series, $X_t = [X_{1t}, \ldots, X_{m,t}]^T$ and $Y_t = [Y_{1t}, \ldots, Y_{mt}]^T$. Let $m = m_1 + m_2$.

Suppose that X_t and Y_t are cointegrated with the residuals ECM_t. The VECM model for testing the Granger causality is given by

$$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} A_x \\ A_y \end{bmatrix} + \begin{bmatrix} A_{xx}(L) & A_{xy}(L) \\ A_{yx}(L) & A_{yy}(L) \end{bmatrix} \begin{bmatrix} \Delta X_{t-1} \\ \Delta Y_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$
$$ECM_{t-1} + \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \qquad (10)$$

where A_x and A_y are m_1 and m_2 dimensional vectors of intercept terms, respectively, $A_{xx}(L)$, $A_{xy}(L)$, $A_{yx}(L)$ and $A_{yy}(L)$ are $n_1 \times n_1$, $n_1 \times n_2$, $n_2 \times n_1$ and $n_2 \times n_2$ dimensional matrices of lag polynomials, respectively, α_x and α_y are n_1 and n_2 dimensional coefficient vectors for the error correction term ECM_{t-1}, respectively. The lag length was selected using the two-stage procedure [42].

There are four different cases of causal relationships between two vectors of time series X_t and Y_t [33].

(1) If $A_{xy}(L)$ is significantly different from the zero, while $A_{yx}(L)$ shows no significantly different from zero, then there exists a unidirectional Ganger causality from time series Y_t to X_t;
(2) If $A_{yx}(L)$ is significantly different from zero, while $A_{xy}(L)$ shows no significantly difference from zero, then there exists a unidirectional Ganger causality from X_t to Y_t;
(3) If both coefficients $A_{xy}(L)$ and $A_{yx}(L)$ are significantly different from zero, then there exists bidirectional Granger causality between X_t and Y_t;
(4) If both coefficients $A_{xy}(L)$ and $A_{yx}(L)$ are not significantly different from zero, then X_t and Y_t are not rejected to be independent.

The four statements imply that Ganger causal relationships between X_t and Y_t depend on the coefficients $A_{xy}(L)$ and $A_{yx}(L)$. Therefore, the null hypotheses for testing the Ganger causality between X_t and Y_t are.

(1) $H_0^1 : A_{xy}(L) = 0$,
(2) $H_0^2 : A_{yx}(L) = 0$, and
(3) Both H_0^1 and $H_0^2 : A_{xy}(L) = 0$ and $A_{yx}(L) = 0$.

Likelihood ratio tests for multivariate Granger causality are given by.

(1) The likelihood ration statistics for testing the null hypothesis: $H_0^1 : A_{xy}(L) = 0$ is

$$T_{ml} = \left(T - (np + 1)\right)\left(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}|\right), \qquad (11)$$

which is asymptotically distributed as a central $\chi^2_{(n_1 \times n_2 \times p)}$ under the hull hypothesis H_0^1.

(2) The likelihood ration statistics for testing the null hypothesis: $H_0^2 : A_{yx}(L) = 0$ is

$$T_{ml} = \left(T - (np + 1)\right)\left(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}|\right),$$

which is asymptotically distributed as a central $\chi^2_{(n_2 \times n_1 \times p)}$ under the hull hypothesis H_0^2.

(3) The likelihood ration statistics for testing the null hypothesis: H_0^1 and $H_0^2 : A_{xy}(L) = 0$ and $A_{yx}(L) = 0$ is

$$T_{ml} = \left(T - (np + 1)\right)\left(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}|\right),$$

which is asymptotically distributed as a central $\chi^2_{(2n_2 \times n_1 \times p)}$ under the hull hypothesis H_0^1 and H_0^2.

The total number of variables to be tested was 18. The p-value for declaring significance after Bonferroni correction was 0.0028.

Data Collection

Data on the number of new cases and new deaths of COVID-19 across the 50 states in the US were obtained from John Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/MAP.HTML). Google mobility indexes were downloaded from Google COVID-19 Community Mobility Reports (https://www.google.com/covid19/mobility/). Comprehensive data and insights on COVID-19's impact on mobility, economy, and society were downloaded from the University of Maryland COVID-19 Impact Analysis Platform (https://data.covid.umd.edu) [43, 44]. All data were collected from March 5, 2020 to August 25, 2020.

RESULTS

Test for Scalar Potential Causes

The scalar variables tested for causation of the new cases and deaths from COVID-19 in the US included the number of contact tracing workers per 100,000 people, percent of population above 60 years of age, median income, population density, percentage of African Americans, percentage of Hispanic Americans, percentage of males, employment density, number of points of interests for crowd gathering per 1,000 people, number of staffed hospital beds per 1,000 people, and number of ICU beds per 1,000 people. The number of new cases and deaths were averaged over time. Each state was a sample. Since the sample sizes were small, the p-value for declaring significance was 0.05 without Bonferroni correction for multiple comparison. The p-values for testing 11 scalar potential causes of the number of new cases and deaths from COVID-19 in the US using both Gaussian kernel and polynomial kernels were summarized in **Table 1** where minimum of three p-values using Gaussian kernel and fourth order and sixth order polynomial kernels were listed and only one causal direction was observed. We observed from **Table 1** that population density (minimum of p-value < 0.0002, which was due to Gaussian kernel), percentage of males (minimum of p-value < 0.03, which was due to Gaussian kernel) and Percentage of Hispanic Americans (minimum of p-value < 0.0325, which was due to sixth order polynomial kernel) showed significant evidence of causing the spread of COVID-19. Employment density (minimum of p-values < 0.0223, which was due to sixth polynomial kernel), Percentage of African American (minimum of p-values < 0.024, which is due to Gaussian kernel), population density (minimum of p-values < 0.025, which was due to Gaussian kernel) and Percentage of males (minimum of p-values <0.0377, which was due to fourth order polynomial) showed significant evidence of causing deaths due to COVID-19. percentage of Hispanic Americans (minimum of

TABLE 1 | *p*-values for testing 10 scalar potential causes of the number of new cases and deaths of COVID-19 in the US.

Risk factor	*p*-value							
	Gaussian kernel		Polynomial (order 6)		Polynomial (order 4)		Minimum	
	New cases	Deaths	New cases	Deaths	New cases	Deaths	New cases	Deaths
Percent of population above the age of 60 years	0.4634	0.1530	0.3472	0.1130	0.2532	0.1547	0.2532	0.1130
Median income	0.1109	0.0760	0.1202	0.0654	0.1713	0.1832	0.1109	0.0654
Percentage of African Americans	0.6526	0.0240	0.9326	0.0710	0.3459	0.0343	0.3459	0.0240
Percentage of Hispanic Americans	0.0575	0.0640	0.0325	0.0520	0.0422	0.1473	0.0325	0.0520
Percentage of males	0.0300	0.1440	0.0403	0.0542	0.0401	0.0377	0.0300	0.0377
Population density	0.0002	0.0250	0.0327	0.0359	0.0265	0.0952	0.0002	0.0250
Employment density	0.4571	0.0590	0.3536	0.0223	0.3485	0.0248	0.3485	0.0223
Number of staffed hospital beds per 1,000 people	0.6732	0.3130	0.5352	0.3224	0.6506	0.2374	0.5352	0.2374
Number of ICU beds per 1,000 people	0.5134	0.4860	0.7516	0.4438	0.6114	0.9037	0.5134	0.4438
Number of contact tracing workers per 100,000 people	0.4203	0.8190	0.4102	0.7215	0.3907	0.7336	0.3907	0.7215

p-value < 0.052, which was due to sixth order polynomial kernel) were close to significance level 0.05 for causing death.

Population density was an important risk factor for both the spread and death from COVID-19. High density resulted in closer contact, stronger interaction among residents and lower social distancing, which facilitated the spread and increased the death rate from COVID-19 [45–48]. However, our results were contradictory with the conclusion of Hamidi et al. [49]. Some literature also confirmed that high proportion of African Americans caused a high rate of deaths [48, 50, 51]. Our results concluded that percentage of Hispanic Americans was a risk factor for the spread and a weak risk factor for death from COVID-19, while the literature showed stronger evidence that Hispanic communities were highly vulnerable to COVID-19 [52].

The second most significant demographic risk factor for the spread of COVID-19 was percentage of males. We found higher COVID-19 morbidity in males than females. However, we did not find higher COVID-19 mortality in males than females.

It was reported that higher COVID-19 mortality in males than females can be due to the following factors [53]. The first factor was higher expression of angiotensin-converting enzyme-2 (ACE two; receptors for coronavirus) in males than females. The second factor was sex-based immunological differences due to sex hormone and the X chromosome.

Test for Granger Causality

Daily mobility and social distancing data from a COVID-19 impacted the analysis platform, including four categories: category A: mobility and social distancing, category B: COVID and health, category C: economic impact, and category D: vulnerable population. A total of 12 temporal metrics in four categories and 12 metrics from the COVID-19 impact analysis platform, six daily Google Community Mobility indexes and protest attendee data that captured real-time trends in movement patterns for each state in the US were included in the analysis to test for Granger causality between these risk factors, health intervention measures and the number of new cases and deaths from COVID-19 across 50 states in the US [44, 54]. The total number of variables to be tested was 18. The *p*-value for declaring significance after Bonferroni correction was 0.0028.

All 18 metrics except for protest attendee showed high significance in causing a reduction of the new cases of COVID-19 in 19 less affected states: VT, WY, ME, AK, NH, WV, ND, SD, NM, RI, DE, KY, KS, CT, CO, IA, WA, WI, and MS. Most of these states were less populated. However, although CA was most affected and the most populated state, all 18 metrics except protest attendance showed a strong significance in causing rapid spread of COVID-19 (**Table 2** and **Supplementary Table S1**). To provide complete causal testing information, we listed the values of statistics for testing 18 temporal potential causes of the number of new cases of COVID-19 across 50 states in the US in **Supplementary Table S2**.

All 18 metrics showed no significance in causing reduction of the new cases of COVID-19 in Florida. The majority of the 18 metrics did not demonstrate evidence that they can significantly mitigate the spread of COVID-19 in most the affected states such as TX, NY, GA, IL, AZ, NJ, NC, and TN. These 10 states were in the top largest states by population in the US. Public health intervention measures such as closing schools and businesses, avoiding public gatherings, restricting traffic, placing residents to stay-at-home and adherence to guidelines were less well implemented or difficult to implement homogeneously due to large populations and geographical areas [55]. These results also explained why the number of new cases of COVID-19 in these states was high and confirmed by several studies [56–58].

Table 3 summarized the ranges of *p*-values, **Supplementary Tables S3, S4** summarized all *p*-values and values of statistics for testing 18 temporal potential causes of the number of new deaths from COVID-19 across 50 states in the US, respectively. All 18 metrics except for protest attendance showed high significant evidence for causing a reduction of new deaths across 50 states except for Michigan (MI) in the US. Our results suggested that a cascade of causes led to the COVID-19 tragedy in the US.

Table 4 listed the most significant risk factor for the new cases of COVID-19 in each of the 50 states in the US. Active Cases/1,000 People, workplaces, number of tests completed/1,000 people, imported COVD cases, unemployment rate and unemployment claims/1,000 people, mobility trends for places of residence (residential), retail and recreation, mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters (retail) and test capacity

TABLE 2 | *p*-values testing 18 temporal potential causes of the number of new cases of COVID-19 in the top 10 most affected states and bottom 10 less affected states in the US.

State	CA	FL	TX	NY	GA	IL	AZ	NJ	NC	TN
Number of cumulative cases	673,095	605,502	586,730	430,774	258,354	224,887	199,273	190,021	157,741	145,417
Miles/person	7.7E-08	3.1E-02	2.3E-03	9.1E-02	2.8E-04	4.9E-04	7.3E-05	4.7E-03	9.2E-06	5.9E-06
Population	1.6E-06	3.6E-02	5.1E-03	6.1E-02	3.2E-04	5.2E-05	1.9E-04	1.2E-03	2.5E-04	2.9E-05
% Change in consumption	1.4E-05	5.5E-02	4.7E-03	1.8E-03	8.7E-04	1.8E-04	2.2E-04	9.6E-07	1.1E-04	1.1E-04
Social distancing index	8.0E-06	9.1E-02	8.8E-03	4.0E-02	3.4E-04	4.6E-04	5.9E-04	1.5E-05	3.5E-03	2.7E-04
Unemployment claims	4.3E-05	6.8E-02	1.7E-02	1.1E-03	2.3E-03	8.0E-04	1.1E-03	2.2E-07	4.2E-03	3.9E-04
Unemployment rate	3.7E-08	2.6E-02	4.4E-03	1.7E-01	4.2E-04	5.3E-06	2.4E-04	2.8E-03	1.5E-04	4.8E-05
% Working from home	5.2E-07	3.7E-02	2.4E-03	3.4E-02	2.4E-04	6.6E-06	2.2E-05	2.8E-04	1.2E-05	2.0E-06
Active cases/1,000 people	2.3E-18	7.2E-02	3.1E-06	5.8E-01	2.8E-10	2.3E-05	7.1E-07	9.5E-04	7.5E-14	3.4E-12
Testing capacity	9.0E-05	7.2E-02	1.3E-02	3.0E-02	2.4E-03	2.1E-05	4.1E-04	1.2E-04	1.7E-03	1.8E-04
Tests done/1,000 people	1.2E-14	3.6E-02	1.1E-05	1.0E-01	2.5E-09	1.5E-03	7.0E-05	3.8E-03	1.3E-07	3.7E-11
Imported COVID cases	1.3E-16	4.5E-01	1.1E-04	2.6E-01	1.7E-07	1.3E-03	5.1E-04	5.9E-04	1.7E-06	4.4E-12
Retail	1.3E-05	9.6E-02	2.2E-02	2.9E-02	4.4E-03	5.9E-04	1.4E-03	3.4E-07	4.5E-03	6.1E-04
Grocery	1.5E-05	1.1E-01	4.1E-03	2.0E-01	5.8E-04	1.5E-05	2.7E-04	9.0E-06	2.6E-04	4.5E-05
Parks	2.4E-08	8.2E-02	6.3E-03	4.1E-02	3.1E-02	6.4E-02	5.6E-03	4.0E-03	8.6E-03	1.7E-05
Transit	5.8E-06	6.4E-02	7.3E-03	2.0E-02	3.3E-04	2.2E-07	3.0E-04	9.2E-06	1.2E-03	3.3E-04
Workplaces	1.5E-05	9.6E-03	2.6E-03	3.5E-03	2.1E-05	5.8E-08	9.3E-06	2.2E-05	1.2E-04	6.1E-05
Residential	7.0E-05	1.7E-02	7.2E-03	9.6E-05	3.6E-04	1.7E-07	6.5E-05	1.1E-06	2.8E-04	1.5E-04
Protest	3.6E-01	9.6E-01	6.6E-01	7.7E-01	8.0E-03	5.1E-01	3.8E-01	1.7E-01	4.1E-01	2.3E-03

State	SD	ND	WV	NH	HI	MT	AK	ME	WY	VT
Number of cumulative cases	11,559	10,229	9,395	7,150	6,984	6,624	5,666	4,368	3,634	1,572
Miles/person	3.7E-08	1.1E-04	1.3E-09	3.8E-14	9.5E-03	5.8E-04	2.8E-05	8.2E-11	2.7E-13	4.8E-14
Population	4.2E-08	5.4E-04	6.6E-08	1.7E-18	2.0E-02	1.2E-03	5.4E-04	1.7E-12	1.7E-11	6.8E-16
% Change in consumption	2.1E-08	1.1E-04	4.6E-10	3.6E-19	3.1E-02	1.4E-04	7.2E-05	1.2E-11	4.2E-10	7.5E-19
Social distancing index	5.7E-05	9.8E-03	5.0E-06	2.8E-19	3.3E-02	6.5E-03	2.4E-03	1.6E-08	1.9E-09	1.6E-19
Unemployment claims	5.0E-06	4.3E-03	1.8E-06	3.6E-16	3.1E-02	3.3E-03	1.4E-03	1.9E-07	1.0E-08	2.1E-32
Unemployment rate	3.8E-08	1.3E-03	2.5E-06	4.8E-26	2.9E-02	2.7E-03	2.8E-04	7.1E-18	8.0E-12	3.8E-14
% Working from home	1.5E-09	1.3E-05	1.5E-12	1.4E-16	2.3E-03	1.1E-04	8.3E-04	1.3E-11	2.4E-15	1.8E-16
Active cases/1,000 people	3.9E-09	3.3E-13	1.5E-16	6.2E-27	1.3E-08	4.8E-09	3.2E-07	6.8E-13	2.1E-21	6.9E-15
Testing capacity	1.1E-07	3.4E-05	5.4E-07	1.3E-22	2.1E-02	3.6E-03	2.0E-03	1.7E-14	6.0E-10	1.5E-23
Tests done/1,000 people	4.5E-07	4.6E-11	2.1E-16	1.9E-12	3.8E-05	1.9E-11	6.3E-09	7.4E-07	1.5E-19	5.7E-13
Imported COVID cases	2.7E-07	6.8E-09	9.6E-13	9.0E-32	7.4E-06	2.7E-08	2.0E-06	1.3E-14	5.4E-19	2.5E-13
Retail	1.3E-07	6.1E-03	4.4E-06	4.8E-19	2.9E-02	2.7E-04	3.1E-04	2.0E-09	5.4E-14	1.4E-20
Grocery	1.3E-04	3.7E-04	1.1E-07	1.7E-22	2.0E-02	2.3E-04	7.5E-04	4.3E-10	3.3E-13	5.6E-20
Parks	3.9E-06	4.9E-04	1.6E-06	3.9E-12	1.9E-02	1.6E-07	2.8E-04	7.8E-07	1.6E-19	3.1E-12
Transit	3.3E-06	3.2E-03	7.4E-09	1.3E-22	2.2E-02	1.6E-04	5.6E-04	6.3E-11	2.0E-13	3.0E-19
Workplaces	3.9E-12	2.3E-04	1.2E-07	1.4E-23	1.8E-02	4.6E-04	4.7E-04	9.8E-15	1.9E-09	2.2E-19
Residential	3.5E-09	1.6E-03	2.0E-06	3.0E-24	1.6E-02	1.4E-03	6.3E-04	4.9E-12	5.6E-08	3.1E-22
Protest	2.4E-01	2.6E-01	9.7E-01	1.1E-03	1.3E-02	2.1E-01	9.6E-03	8.0E-02	8.3E-01	1.6E-01

Retail: Retail and recreation, mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.

Grocery: Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.

Parks: Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens.

Transit: Transit stations, mobility trends for places like public transport hubs such as subway, bus, and train stations.

Workplaces: Mobility trends for places of work.

Residential: Mobility trends for places of residence.

Test Rate: Ratio of the number of individuals who have taken the virus test over the total population in the region.

Attendee: Number of attendees in the protest.

were the most significant risk factors for the new cases of COVID-19 in 23, 7, 6, 5, 4, 2, 1 and 1 states of the US, respectively.

Table 5 summarized the most significant risk factor for the deaths from COVID-19 in each of the 50 states in the US. Active Cases/1,000 people, workplaces, residential, unemployment rate, imported COVID cases, unemployment claims/1,000 people, transit, test done/1,000 people, grocery, testing capacity, retail, percentage of change in consumption, percentage of working from home were the most significant risk factor for the deaths of COVID-19 in 17, 10, 4, 4, 3, 2, 2, 2, 1, 1, 1, 1 states, respectively. We also observed that the number of protest attendees showed

mild significant evidence to cause increasing the number of new cases of COVID-19 in KY (*p*-value < 0.00012), KS (*p*-value < 0.00026), NH (*p*-value < 0.00108), MA (*p*-value < 0.0016) and TN (*p*-value < 0.0024) or to cause more deaths from COVID-19 in OR (*p*-value < 5.11 E-05), TX (*p*-value < 0.00017), ME (*p*-value < 0.00028), KS (*p*-value < 0.00061), MI (*p*-value < 0.0015), OH (*p*-value < 0.0021) and NC (*p*-value < 0.0023).

In some cases, two causal directions may occur. To examined whether the COVID-19 cases and deaths caused daily mobility and social distancing metrics to change, we summarized the values of statistics and *p*-values for testing COVID-19 new case

TABLE 3 | Ranges of *p*-values for testing 18 temporal potential causes of the number of new deaths from COVID-19 across 50 states in the US.

State	p-value		State	p-value	
	Lower bound	Upper bound		Lower bound	Upper bound
AK	1.47E-18	3.95E-24	MT	7.66E-15	4.41E-25
AL	7.45E-11	1.10E-18	NC	2.93E-08	8.95E-18
AR	1.17E-12	1.54E-30	ND	2.54E-17	1.04E-20
AZ	4.35E-11	1.29E-27	NE	5.11E-15	2.62E-23
CA	1.10E-04	1.54E-13	NH	6.67E-15	3.41E-28
CO	1.41E-21	6.44E-26	NJ	1.80E-03	1.96E-09
CT	6.73E-07	3.13E-16	NM	7.20E-07	1.15E-16
DE	1.20E-03	1.35E-06	NV	9.00E-09	1.45E-17
FL	2.50E-04	1.47E-15	NY	4.30E-02	1.52E-05
GA	6.53E-07	5.81E-17	OH	1.82E-08	2.67E-17
HI	1.77E-15	1.14E-20	OK	4.29E-07	1.89E-19
IA	6.78E-06	8.06E-15	OR	3.37E-15	4.50E-26
ID	1.96E-08	7.11E-18	PA	3.86E-15	1.11E-23
IL	3.81E-05	1.34E-12	RI	4.26E-09	2.16E-22
IN	1.78E-07	3.39E-17	SC	8.97E-08	6.75E-25
KS	8.74E-30	1.42E-40	SD	5.04E-18	6.01E-25
KY	1.89E-13	1.40E-24	TN	7.28E-09	2.06E-24
LA	7.71E-11	2.38E-20	TX	6.22E-07	2.59E-23
MA	1.25E-03	3.92E-04	UT	3.90E-13	8.12E-29
MD	3.80E-02	7.63E-06	VA	3.39E-17	2.61E-16
ME	2.29E-15	4.29E-25	VT	3.39E-07	1.23E-29
MI	0.28	2.07E-05	WA	5.67E-16	3.83E-29
MN	2.40E-04	8.89E-11	WI	2.21E-10	4.77E-29
MO	1.73E-11	7.85E-20	WV	9.61E-15	7.81E-22
MS	2.94E-08	1.49E-19	WY	7.93E-26	2.39E-28

and death potential causes of six Google mobility indexes across 50 states in the US in **Supplementary Tables S5–7** and DS8,

respectively. We did observe the opposite causal direction. **Table 6** presented the smallest *p*-value across 50 states in the US for testing two causal directions: from Google mobility indexes to the number of new cases of COVID-19 and from the number of new cases of COVID-19 to the Google mobility indexes. We observed significant causation from the number of COVID-19 to the Google mobility indexes in some states. However, the *p*-values for testing causation from the number of new cases of COVID-19 to the Google mobility indexes were much larger than that from the Google mobility indexes to the number of new cases of COVID-19. The causal pattern for the COVID-19 deaths was similar to the number of new cases of COVID-19.

To illustrate the causal relationships between the risk factors and the number of new cases and deaths from COVID-19, we plotted **Figures 1** and **2**. **Figure 1** plotted the social distance index curves as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL) and Rhode Island (RI). **Figure 1** showed that the social distance index in FL was much higher than that in RI state, which resulted in the larger number of new cases of COVID-19 in FL than that in RI. **Figure 2** showed the number of imported COVID-19 cases as a function of time from March 5, 2020 to August 25, 2020 in Maryland (MD) and Wyoming (WY). We observed a huge difference in the number of imported COVID-19 cases between MD and RI. The very low number of imported cases of COVID-19 in WY resulted in the very low number of deaths from COVID-19 in WY, while the high number of imported cases in MD state led to the increased deaths from COVID-19 in MD. These results were consistent with the finding in the literature. It was reported that strong

TABLE 4 | The most significant risk factor for the new cases of COVID-19 in each of 50 states in the US.

State	Risk factor	p-value	State	Risk factor	p-value
AK	Tests done/1,000 people	6.27E-09	MT	Tests done/1,000 people	1.87E-11
AL	Active cases/1,000 people	5.40E-08	NC	Active cases/1,000 people	7.49E-14
AR	Active cases/1,000 people	4.49E-14	ND	Active cases/1,000 people	3.35E-13
AZ	Active cases/1,000 people	7.12E-07	NE	Unemployment rate	6.10E-10
CA	Active cases/1,000 people	2.29E-18	NH	Imported COVID cases	9.04E-32
CO	Work places	1.48E-17	NJ	Unemployment claims/1,000 people	2.15E-07
CT	Test capacity	5.46E-25	NM	Active cases/1,000 people	5.68E-13
DE	Residential	3.64E-15	NV	Imported COVD cases	2.36E-09
FL	Work places	9.60E-03	NY	Residential	9.57E-05
GA	Active cases/1,000 people	2.76E-10	OH	Active cases/1,000 people	6.42E-10
HI	Active cases/1,000 people	1.31E-08	OK	Tests done/1,000 people	4.14E-10
IA	Active cases/1,000 people	5.07E-14	OR	Active cases/1,000 people	3.87E-12
ID	Imported COVD cases	4.53E-10	PA	Work places	4.21E-09
IL	Work places	5.82E-08	RI	Imported COVD cases	1.02E-27
IN	Active cases/1,000 people	1.12E-09	SC	Active cases/1,000 people	1.58E-04
KS	Tests done/1,000 people	1.21E-32	SD	Work places	3.89E-12
KY	Imported COVID cases	4.30E-26	TN	Active cases/1,000 people	3.45E-12
LA	% Working from Home	1.58E-19	TX	Active cases/1,000 people	3.05E-06
MA	Retail	3.20E-10	UT	Active cases/1,000 people	2.04E-07
MD	Work places	6.35W-09	VA	Active cases/1,000 people	2.08E-08
ME	Unemployment rate	7.15E-18	VT	Unemployment claims/1,000 people	2.07E-32
MI	Work places	1.16E-10	WA		2.65E-21
MN	Active cases/1,000 people	9.01E-07	WI	Active cases/1,000 people	5.19E-12
MO	Tests done/1,000 people	2.42E-09	WV	Active cases/1,000 people	1.53E-16
MS	Tests done/1,000 people	2.16E-12	WY	Active cases/1,000 people	2.13E-21

TABLE 5 | The most significant risk factor for deaths of COVID-19 in each of 50 states in the US.

State	Risk factor	p-value	State	Risk factor	p-value
AK	Active cases/1,000 people	3.95E-24	MT	Active cases/1,000 people	4.41E-25
AL	Active cases/1,000 people	1.10E-18	NC	Work places	8.95E-18
AR	Tests Done/1,000 people	1.54E-30	ND	Unemployment rate	1.04E-20
AZ	Active cases/1,000 people	1.29E-27	NE	Work places	2.62E-23
CA	Work places	1.54E-13	NH	Imported COVID cases	3.41E-28
CO	Residential	6.44E-26	NJ	Residential	1.96E-09
CT	Active cases/1,000 people	3.12E-16	NM	Unemployment rate	1.15E-16
DE	Retail	1.35E-06	NV	Active cases/1,000 people	1.45E-17
FL	Active cases/1,000 people	1.47E-15	NY	% Change in Consumption	1.52E-05
GA	Work places	5.81E-17	OH	Unemployment rate	2.67E-17
HI	Active cases/1,000 people	1.14E-20	OK	Work places	1.89E-19
IA	Unemployment rate	8.06E-15	OR	% Working from Home	4.50E-26
ID	Active cases/1,000 people	7.11E-18	PA	Imported COVID cases	1.11E-23
IL	Residential	1.34E-12	RI	Imported COVID cases	2.16E-22
IN	Work places	3.39E-17	SC	Active cases/1,000 people	6.75E-25
KS	Grocery	1.42E-40	SD	Active cases/1,000 people	6.01E-24
KY	Work places	1.40E-24	TN	Test Done/1,000 people	2.06E-24
LA	Transit	2.38E-20	TX	Active cases/1,000 people	2.59E-23
MA	Active cases/1,000 people	3.92E-09	UT	Active cases/1,000 people	8.12E-29
MD	Grocery	7.60E-06	VA	Work places	2.61E-16
ME	Residential	4.29E-25	VT	Unemployment claims/1,000 people	1.23E-29
MI	Unemployment claims/1,000 people	2.07E-05	WA	Transit	3.83E-29
MN	Testing capacity	8.89E-11	WI	Work places	4.77E-29
MD	Work places	7.85E-20	WV	Active cases/1,000 people	7.81E-22
MS	Active cases/1,000 people	1.49E-19	WY	Active cases/1,000 people	2.39E-28

TABLE 6 | Lower bound of p-values for testing causation from Google mobility indexes to COVID-19 case and from COVID-19 case to Google mobility indexes.

Cause	Effect	Smallest p-value	Cause	Effect	Smallest p-value
Retail	COVID-19 case	7.67E-26	COVID-19 case	Retail	1.29E-04
Grocery		3.70E-25		Grocery	4.85E-11
Parks		3.87E-23		Parks	1.80E-12
Transit		1.90E-27		Transit	1.29E-04
Workplaces		1.07E-25		Workplaces	3.99E-06
Residential		1.09E-25		Residential	2.99E-07

interventions would substantially decrease the number of deaths (Davies et al. 2020; Gagnon et al. 2020).
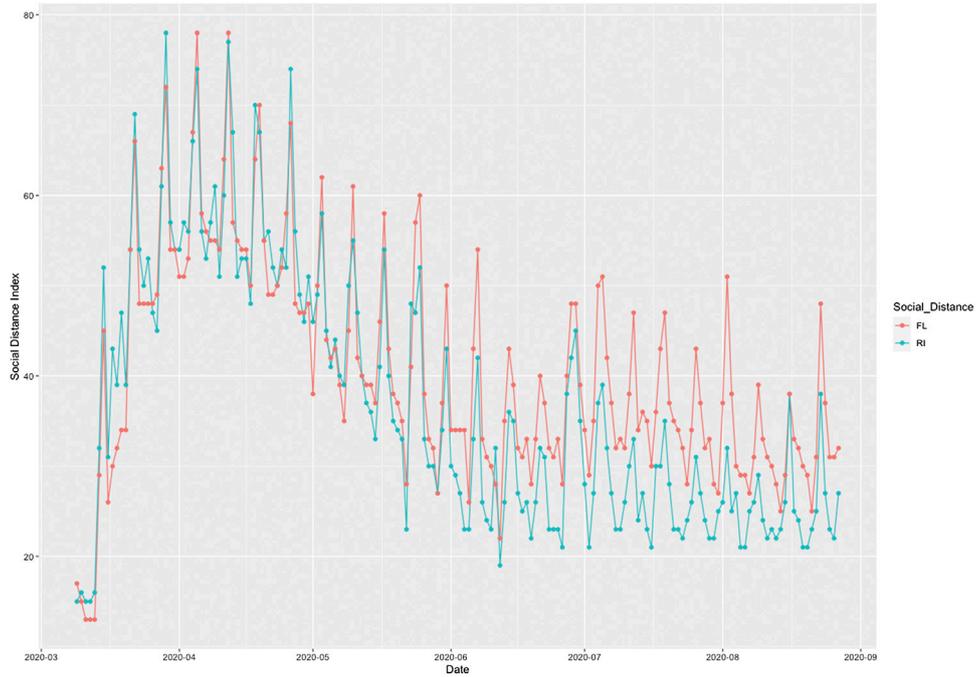
The COVID-19 case curves as a function of time from March 5, 2020 to August 25, 2020 in FL, RI, MD and WY were plotted in **Figure 3**, respectively. The response of COVID-19 to public health intervention which were measured by Google mobility indexes were usually delayed.

DISCUSSION

Causal inference for COVID-19 is essential for selecting and implementing public intervention measures and understanding the role of the demographics in curbing the spread and reducing the deaths from COVID-19. In this paper, we systematically addressed the issues in identifying causal risk factors and evaluating the causal effects of risk factors and intervention measures on the spread and deaths from COVID-19 in the US. Risk factors and intervention measures included

scalar variables and temporal variables. The ANMs were used to test for causal relationships between scalar risk factors and the average number of new cases or deaths from COVID-19 in the US. Transmission of COVID-19 is a dynamic system. Many risk factors and intervention measures are temporal variables. The Granger Causality Test was used to reveal the causal relationships between the temporal risk factors and intervention measures, and the number of new cases or new deaths from COVID-19 across the 50 states in the US.

The demographic risk factors were the major part of the scalar risk factors in the causal analysis of COVID-19. We found that population density was the most significant causal factor of both new cases and death from COVID-19. Population density measured the average number of people per square kilometer living in a built-up area. Densely populated states generated conditions where COVID-19 can spread quickly and undetected in the densely populated areas and created high levels of vulnerability. The second significant demographic factor was percentage of males. Our data

FIGURE 1 | Social distance index curves as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL) (red color) and Rhode Island (RI) (blue color).
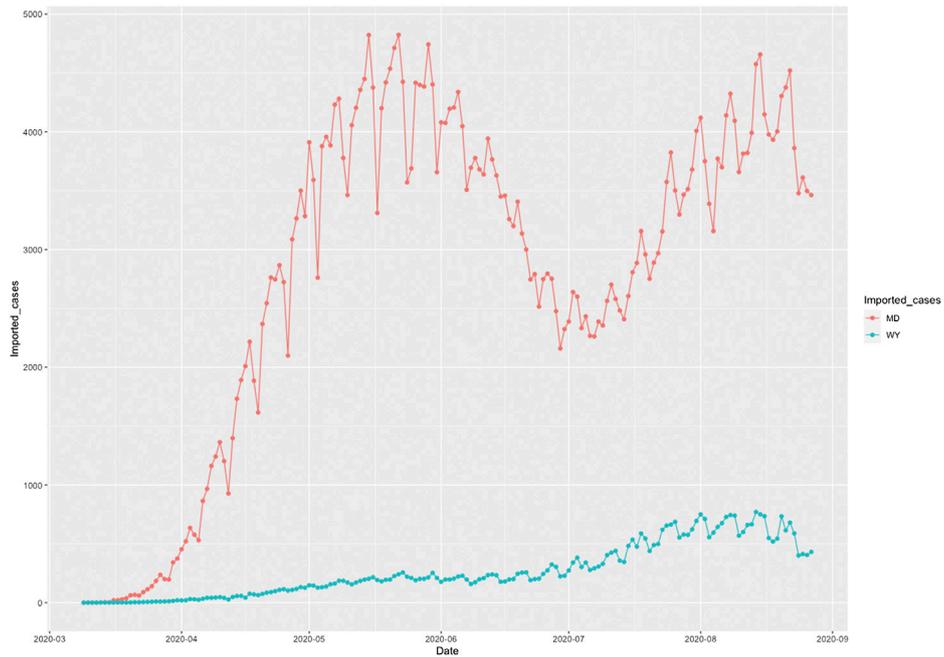
FIGURE 2 | Number of imported COVID-19 cases as a function of time from March 5, 2020 to August 25, 2020 in Maryland (MD) (red color) and Wyoming (WY) (blue color).

FIGURE 3 | Number of new cases of COVID-19 as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL), Rhode Island (RI), Maryland (MD), and Wyoming (WY).
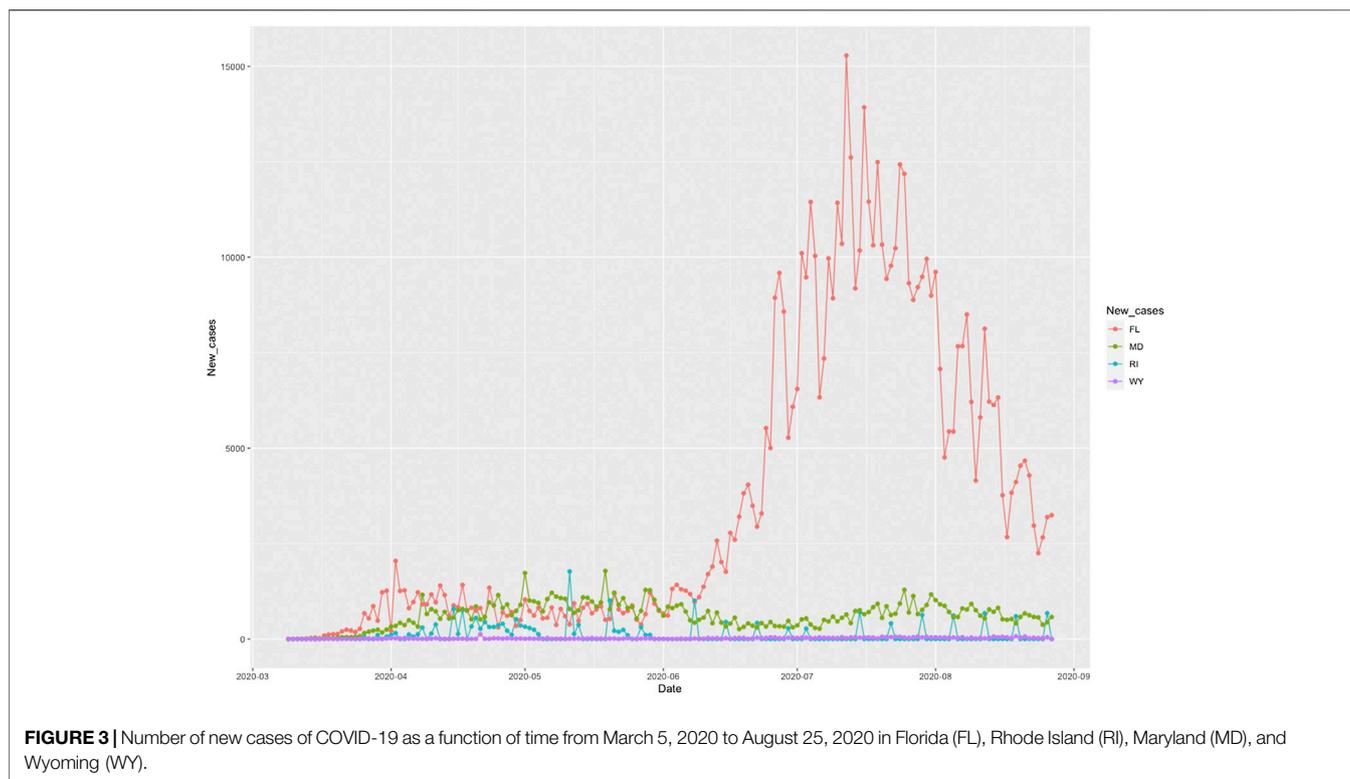
suggested that men were more vulnerable to COVID-19 than women. However, our analysis did not conclude that more men than women were dying from COVID-19.

We also discovered that more Black Americans were dying from COVID-19. The reasons for this were complex. Black Americans had higher rates of chronic disease conditions, including diabetes, heart disease, and lung disease, were poor and more easily exposed to the COVID-19, and lived in the cramped housing. Inequities in the social determinants of health affected mortality and morbidity of COVID-19 for Hispanic Americans with much milder significance.

We studied the causal effect of major public health interventions across the 50 states in the US. In the absence of centralized intervention measures and implementation of a timeline and presence of the complex dynamics of human mobility and the variable intensity of local outbreaks of COVID-19, evaluating the causal effect of public health intervention measures on COVID-19 transmission and deaths in the USA posed a great challenge. We used 6 Google mobility indexes and 12 daily metrics to measure the effects of COVID-19 spread and public health interventions on mobility and social distancing, derived from mobile device location data and COVID-19 case data, provided by the University of Maryland COVID-19 Impact Analysis Platform. These real time metrics capture the dynamics of social distancing. Granger causality tests were used to identify the causality relationships between time series metrics and time varying in the number of new cases or deaths from COVID-19. Although the risk factors differed by location, Active Cases/1,000 people were a significant risk factor

for both number of new cases and deaths from COVID-19 in most states. The most popular intervention measure in the US was workplaces (mobility trends for places of work). Workplaces were the significant cause of the number new cases of COVID-19 in 44 states and significant cause of death in 49 states. Therefore, workplaces should be considered as a very important risk mitigation measure to reduce the number of new cases and deaths from COVID-19. Tests done/1,000 people was the second population intervention in the US. It was the significant cause of the new cases of COVID-19 in 46 states and significant cause of death in 47 states. Virus test results in quick case identification and isolation to contain COVID-19, and rapid treatment to reduce the number of deaths. Imported COVID cases were also a top significant risk factor for speeding the spread and increasing the deaths from COVID-19. Our results showed that the imported COVID case metric was the significant causal factor for the new cases in 46 states and the significant causal factor for the deaths in 47 states.

Our results showed that the high numbers of cases and deaths from COVID-19 were due to lacking strong interventions and high population density. We observed that no metrics showed significant evidence in mitigating the COVID-19 epidemic in FL and only a few metrics showed evidence in reducing the number of new cases of COVID-19 in AZ, NY and TX. Our results showed strong interventions were needed to contain COVID-19.

Although we tried to systematically and comprehensively analyze the data, this study has multiple limitations. First, we only analyzed the causal relationship between mobility patterns and the number of new cases or deaths and ignored the role of

other potential mitigating factors (e.g., wearing face masks) that could also have contributed to the reduction of new cases or deaths from COVID-19. When data are available, more metrics should be included in the analysis.

Second, we have not addressed the confounding bias issue. When confounding is unknown, adjusting for confounding methods cannot be applied to eliminate confounding bias from the causal analysis. Unadjusted confounding bias will distort the inferred (true) causal relationship between the number of new cases or deaths from COVID-19, and metrics for social distancing when these two variables share common causes. This will have substantive implications for developing interventions to mitigate the spread of COVID-19 and reduce the deaths from COVID-19. However, removing confounding from causal analysis for COVID-19 is complicated and will be investigated in the future.

In summary, our analysis has provided information for both individuals and governments to plan future interventions on containing COVID-19 and reduction of deaths from COVID-19.

DATA AVAILABLITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZL, data analysis; TX, data analysis; KZ, data acquisition and interpretation; H-WD, interpretation; EB, biological interpretation; MX, concept design, method development and write a manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2020.611805/full#supplementary-material.

REFERENCES

1. Callaway E. Time to use the p-word? Coronavirus enters dangerous new phase. *Nature* (2020) doi:10.1038/d41586-020-00551-1

2. Priyadarsini SL, Suresh M. Factors influencing the epidemiological characteristics of pandemic COVID 19: a TISM approach. *Int J Healthc Manag* (2020) 13:89–98. doi:10.1080/20479700.2020.1755804

3. Irfan U. The math behind why we need social distancing, starting right now. (2020) Available from: https://www.vox.com/2020/3/15/21180342/coronavirus-covid-19-us-social-distancing (Accessed March 15, 2020)

4. Farseev A, Chu-Farseevac YY, Yanga Q, Loo DB. Understanding economic and health factors impacting the spread of COVID-19 disease. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.10.20058222 (Accessed June 9, 2020)

5. Tantrakarnapa K, Bhopdhornangkul B, Nakhaapakorn K. Influencing factors of COVID-19 spreading: a case study of Thailand. *J Public Health* (2020) [Epub ahead of print]. doi:10.1007/s10389-020-01329-5

6. Nakada LYK, Urban RC. COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in the expanded metropolitan area of São Paulo, Brazil. *Environ Sci Pollut Res Int* (2020) 28:1–7. doi:10.1007/s11356-020-10930-w

7. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *WClinical Medicine* (2020) 25:100464. doi:10.1016/j.eclinm.2020.100464

8. Baum CF, Henry M. Socioeconomic factors influencing the spatial spread of COVID-19 in the United States. In: *Boston college working papers in economics*, 1009. Boston College Department of Economics (2020)

9. Coccia M. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci Total Environ* (2020) 729:138474. doi:10.1016/j.scitotenv.2020.138474

10. Livadiotis G. Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19. *PloS One* (2020) 15(5):e0233875. doi:10.1371/journal.pone.0233875

11. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* (2020) 395:1054–1062. doi:10.1016/S0140-6736(20)30566-3

12. Raghupathi V. An empirical investigation of chronic diseases: a visualization approach to Medicare in the United States. *Int J Healthc Manag* (2019) 12(4):327–39. doi:10.1080/20479700.2018.1472849

13. Anderson RM, Hollingsworth TD, Baggaley RF, Maddren R, Vegvari C. COVID-19 spread in the UK: the end of the beginning? *Lancet* (2020) 396(10251):587–90. doi:10.1016/S0140-6736(20)31689-5

14. Saadat S, Rawtani D, Hussain CM. Environmental perspective of COVID-19. *Sci Total Environ* (2020) 728(2020):138870. doi:10.1016/j.scitotenv.2020.138870

15. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020) 584:257–61. doi:10.1038/s41586-020-2405-7

16. MonodUnwin RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc Health* (2020) 4:397–404. doi:10.1016/S2352-4642(20)30095-X

17. BooyMytton J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. *Nat Med* (2020) 26:1183–92. doi:10.1038/s41591-020-1011-4

18. McKendryRees CN, Iboi E, Eikenberry S, Scotch M, MacIntyre CR, Bonds MH, et al. Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. *Math Biosci* (2020) 325:108364. doi:10.1016/j.mbs.2020.108364

19. Gumel SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. *JMIR Pediatr Parent* (2020) 3(1):e20049. doi:10.2196/20049

20. Altman N, Krzywinski M Association, correlation and causation. *Nat Methods* (2015) 12(10):899–900. doi:10.1038/nmeth.3587

21. Sharkey P, Wood G. The causal effect of social distancing on the spread of SARS-CoV-2 (2020) Available from: https://doi.org/10.31235/osf.io/hzj7a (Accessed May 19, 2020)

22. Steigera E, Mußgnuga T, Kroll LE. Causal analysis of COVID-19 observational data in German districts reveals effects of mobility, awareness, and temperature. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.07.15.20154476 (Accessed July 23, 2020)

23. Li AY, Hannah TC, Durbin J, Dreher N, McAuley FM, Marayati NF, et al. Multivariate analysis of factors affecting COVID-19 case and death rate in U.S counties: the significant effects of Black race and temperature. *Am Med J Sci* (2020) 360(4):348–56. doi:10.1016/j.amjms.2020.06.015

24. Zenil H, Kiani NA, Zea AA, Tegnér J. Causal deconvolution by algorithmic generative models. *Nat Mach Intell* (2019) 1:58–66. doi:10.18148/srm/2020.v14i2.7723

25. Ramachandra V, Sun H. Causal inference for COVID-19 interventions. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.09.29.20203505 (Accessed September 29, 2020)

26. Friston KJ, Parr T, Zeidman P, Razi A, Flandin G, Daunizeau J, et al. Dynamic causal modelling of COVID-19. *Wellcome Open Res* (2020) 5:89. doi:10.12688/wellcomeopenres.15881.2

27. LambertHulme V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early COVID-19 pandemic in the U.S. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.05.27.20115139 (Accessed September 12, 2020)

28. Regis Annea W, Carolin Jeeva S. ARIMA modelling of predicting COVID-19 infections. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.18.20070631 (Accessed April 23, 2020)

29. Peters J, Mooij JM, Janzing D, Schölkopf B. Causal discovery with continuous additive noise models. *J Mach Learn Res* (2014) 15:2009–53. 10.15496/publikation-1672

30. Johansen S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* (1991) 59(6):1551–80. doi:10.2307/2938278

31. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* (1969) 37:424–38. doi:10.2307/1912791

32. Eichler M. Causal inference with multiple time series: principles and problems. *Philos Trans A Math Phys Eng Sci* (2013) 371:20110613. doi:10.1098/rsta.2011.0613

33. Bai ZD, Wong WK, Zhang BZ. Multivariate linear and nonlinear causality tests. *Math Comput Simulat* (2010) 81:5–17. doi:10.1371/journal.pone.0185155

34. Jiao R, Lin N, Hu Z, Bennett DA, Jin L, Xiong M. Bivariate causal discovery and its applications to gene expression and imaging data analysis. *Front Genet* (2018) 9:347. doi:10.3389/fgene.2018.00347

35. Mooij JM, Peters J, Janzing D, Zscheischler J, Schölkopf B. Distinguishing cause from effect using observational data: methods and benchmarks. *J Mach Learn Res* (2016) 17(32):1–102.

36. Heydari MR, Salehkaleybar S, Zhang K. Adversarial orthogonal regression: two non-linear regressions for causal inference (2019) Available at: https://arxiv.org/abs/1909.04454 (Accessed September 10, 2019)

37. Wang Y. *Smoohing splines: methods and applications.* 1st ed. Chapman and Hall/CRC (2011)

38. Lu S, Shen Y, Wang Y. Generalized high-precision simulation for TT&C channels using B-spline signal processing. *IEEE Signal Process Lett* (2017) 24(9):1383–7. doi:10.1109/LSP.2017.2727524

39. Cleveland WS. Robust locally weighted regression and smoothing scatterplot. *J Am Stat Assoc* (1979) 74(368):829–36. 10.1080/01621459.1979.10481038

40. Gretton A, Bousquet O, Smola A, Schölkopf B. Measuring statistical dependence with Hilbert–Schmidt norms. Proceedings of the International Conference on Algorithmic Learning Theory; 2005 October 8–11; Singapore (2005) 63–77.

41. Vakhania NN, Tarieladze VI, Chobanyan SA. Covariance operators. In: *Probability distributions on banach spaces. Mathematics and its application (soviet series)*, Vol. 14. Dordrecht: Springer (1987)

42. Abdalla I, Murinde V. Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan and the Philippines. *Appl Financ Econ* (1997) 7:25–35. doi:10.1080/096031097333826

43. Maryland Transportation Institute. *University of Maryland COVID-19 impact analysis platform.* College Park, USA: University of Maryland (2020) Available from: https://data.covid.umd.edu (Accessed 2020)

44. Zhang L, Ghader S, Pack M, Darzi A, Xiong C, Yang M, et al. *An interactive COVID-19 mobility impact and social distancing analysis platform.* medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.29.20085472 (Accessed May 5, 2020)

45. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *J Trav Med* (2020) 27(3):taaa038. doi:10.1093/jtm/taaa038

46. Pequeno P, Mendel B, Rosa C, Bosholn M, Souza JL, Baccaro F, et al. Air transportation, population density and temperature predict the spread of COVID-19 in Brazil. *PeerJ* (2020) 8:e9322. doi:10.7717/peerj.9322

47. MagnussonBarbosa E. High population density in India associated with spread of COVID-19. Medical Research News (2020) Available from: https://www.news-medical.net/news/20200703/High-population-density-in-India-associated-with-spread-of-COVID-19.aspx (Accessed July 3, 2020)

48. Rajan K, Dhana K, Barnes LL, Aggarwal NT, Evans L, Wilson RS, et al. Strong effects of population density and social characteristics on distribution of COVID-19 infections in the United States. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.05.08.20073239 (Accessed May 19, 2020)

49. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? *J Am Plann Assoc* (2020) 86:495–509. doi:10.1080/01944363.2020.1777891

50. Golestaneh L, Neugarten J, Fisher M, Billett HH, Gil MR, Johns T, et al. The association of race and COVID-19 mortality. *EClinicalMedicine* (2020) 25:100455. doi:10.1016/j.eclinm.2020.100455

51. BellinYunes UV, Larkins-Pettigrew M. Racial demographics and COVID-19 confirmed cases and deaths: a correlational analysis of 2886 US counties. *J Public Health* (2020) 42(3):445–7. doi:10.1093/pubmed/fdaa070

52. Calo WA, Murray A, Francis E, Bermudez M, Kraschnewski J. Reaching the hispanic community about COVID-19 through existing chronic disease prevention programs. *Prev Chronic Dis* (2020) 17:200165. doi:10.5888/pcd17.200165

53. Bwire GM. Coronavirus: why men are more vulnerable to covid-19 than women?. *SN Compr Clin Med* (2020) 4:1–3. doi:10.1007/s42399-020-00341-w

54. *Google community mobility reports.* (2020) Available from: https://www.google.com/covid19/mobility/ (Accessed August 17, 2020)

55. Althouse BM, Wallace B, Case B, Scarpino SV, Berdhal A, White ER, et al. The unintended consequences of inconsistent pandemic control policies. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.08.21.20179473 (Accessed October 28, 2020)

56. Hebert-Dufresne S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.03.03.20029843 (Accessed March 13, 2020)

57. Masrur A, Yu M, Luo W, Dewan A. Space-time patterns, change, and propagation of COVID-19 risk relative to the intervention scenarios in Bangladesh. *Int J Environ Res Publ Health* (2020) 17(16):E5911. doi:10.3390/ijerph17165911

58. Goldschmidt-Clermont PJ. COVID-19 real-world data for the US and lessons to reopen business. *PLoS Pathog* (2020) 16(8):e1008756. doi:10.1371/journal.ppat.1008756