scientific reports

Check for updates

OPEN

Causal connections between socioeconomic disparities and COVID-19 in the USA

Tannista Banerjee[1], Ayan Paul[2,3]✉, Vishak Srikanth[4,5] & Inga Strümke[6,7]

With the increasing use of machine learning models in computational socioeconomics, the development of methods for explaining these models and understanding the causal connections is gradually gaining importance. In this work, we advocate the use of an explanatory framework from cooperative game theory augmented with *do* calculus, namely causal Shapley values. Using causal Shapley values, we analyze socioeconomic disparities that have a causal link to the spread of COVID-19 in the USA. We study several phases of the disease spread to show how the causal connections change over time. We perform a causal analysis using random effects models and discuss the correspondence between the two methods to verify our results. We show the distinct advantages a non-linear machine learning models have over linear models when performing a multivariate analysis, especially since the machine learning models can map out non-linear correlations in the data. In addition, the causal Shapley values allow for including the causal structure in the variable importance computed for the machine learning model.

The early stages of the spread of COVID-19 in the USA laid bare how socioeconomic disparities bring about a disproportionate spread of the disease in certain parts of society[1–14]. This trend is a pattern that has been historically observed for several diseases caused by viruses like HIV[15], MERS-CoV, SARS-CoV, Ebola[16–18] etc. Even beyond variances in socioeconomic conditions, ethnicity is a contributing factor[19–22] for disparities in healthcare. A good understanding of the causal connection between socioeconomic disparities and the spread of a disease is necessary for implementing policies to mitigate disease spread amongst those who are the most vulnerable.

In a recent work, it was shown that the spread of COVID-19 is correlated to various socioeconomic metrics averaged at the county level[23] during the initial stages of the pandemic. The use of Shapley values was introduced with a machine learning framework to understand how socioeconomic disparities affect the spread of the disease. However, Shapley values decomposing a machine learning model probe the correlations between the endogenous and exogenous variables and do not take the causal connection in the data into account. In this work, we go a step further and introduce causal Shapley values to the analysis of disease spread.

We obtain a regression model by training a machine learning model in a supervised manner on COVID-19 prevalence data. As the resulting model is too complex to be considered interpretable, we use methods from the field of explainable AI to explain the model's predictions. Specifically, we use a framework built on the game-theoretic solution concept of Shapley values[24], upon which the packages of Shapley additive explanations[25] and shapr[26] are based. The idea behind these was recently built upon by Ref.[27] to avoid independence assumptions in the data as well as to incorporate causal knowledge from causal chain graphs in the Shapley value calculation. Our objectives in this work can be delineated as:

- We study the spread of COVID-19 in three different regions of the USA broken down by the pattern of spread of the disease.
- We build socioeconomic metrics using data sources from the US Census Bureau at the county level to model the demographic distribution of the population.
- We use an ensemble of boosted decisions trees (BDTs) to build a regression model of the data.
- We calculate the causal Shapley values using these regression models to understand the causal connections between the socioeconomic metrics and the spread of COVID-19.

[1]Department of Economics, Auburn University, 140 Miller Hall, Auburn, AL 36849, USA. [2]DESY, Notkestraße 85, 22607 Hamburg, Germany. [3]Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany. [4]BASIS Independent Silicon Valley, San Jose, CA, USA. [5]Stanford Online High School, Stanford, CA, USA. [6]Department of Engineering Cybernetics, NTNU, 7034 Trondheim, Norway. [7]Department of Holistic Systems, SimulaMet, 0167 Oslo, Norway. ✉email: ayan.paul@desy.de

- We study two different times, the early stage the disease spread between February and July 2020 and the later stage of the disease spread between July 2020 and January 2021 to understand the variance in the causal connections.
- We compare the multivariate methods that we establish here with linear random effects models and discuss the relative advantage of using machine learning for highly correlated datasets.

To avoid the effects of causation being washed out due to large scale averaging[23], we divide our analysis into three regions. This is particularly important since the demographic nature of various parts of the USA are different. In addition, the pattern of disease spread were also observed to vary in different parts of the USA. The states with higher densities were significantly affected in the first wave of the pandemic while the southern states were significantly affected by the second wave. The states on the west coast were affected throughout the extent of the pandemic. Demographically, the northeastern states have several urban areas that are closely clustered together. However, in the southern states the urban areas are spread far apart and a large fraction of the population lives in rural areas where the socioeconomic conditions are different from the urban areas. The states on the west coast are a mix of urban and rural areas with a very different economic structure. Hence, we will focus on three different clusters of states (the correlation matrices for all counties in the USA along with those for the three regions can be found in the Supplementary Information section).

- High population density regions: States with population density over 400 individuals per sq. km: District of Columbia, New Jersey, Rhode Island, Massachusetts, Connecticut, Maryland, Delaware and New York.
- The southern states: These include: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia.
- The west coast: These include: California, Oregon and Washington.

Some notable recent works have addressed causal analyses in several distinct directions for COVID-19[28–30]. Our work complements these approaches and provides a well delineated prescription for performing a causal analysis in computational socioeconomics. We hope that by highlighting how the socioeconomic disparities that aggravated the spread of COVID-19 in of the community, and often amongst those who are at a social disadvantage, we can not only aid in future policy building but also establish a robust method for causal analysis that can be easily used for highly multivariate datasets.

Causal ordering

The framework advocated by Heskes et al.[27] for causal Shapley values accounts for indirect effects in order to take the causal structure of the data into account in when estimating the Shapley values. The approach recognizes that it is not practical to compute interventional probabilities to account for indirect effects and instead proposes a causal chain graph The (partial) causal ordering in the data is represented as a causal chain graph by a directed acyclic graph (DAG). A causal ordering (CO) is then a nested list that maps to a sequence of the respective causal orderings of the socioeconomic metrics. Variables of equivalent causal importance are grouped into the same sub-list within the list. For example, [NW, [SC, Emp]] would represent non-white as the most significant cause which influences the next set of variables, here Senior Citizen and Employed, each with equal importance (the abbreviations and definitions of the metrics can be found in the "Methods" section). More intuitively, the causal ordering depends on a hypothesis of how the socioeconomic metrics form subgroups that are causally connected. Within a subgroup, the metrics are assumed to have no causal connection but can be cyclically connected or share confounding variables. There is no fixed prescription for constructing these partial causal orderings. Rather, domain knowledge has to be used to assume how each metric can affect the others to build causal orderings. We investigate six different causal orderings, CO#1–CO#3 described below and CO#4 - CO#6 included in the Supplementary Information, and explain the rationale behind each choice. The DAG that represent the causal orderings are shown in Fig. 1 for CO#1–CO#3.

CO#1. The first causal ordering we construct is [NW, [SC, Emp, Uemp], [Inc, Lab], [Uins, Com, Pov, GI],[Den, MC, Tran]]. The rationale for this baseline ordering is as follows.

- Areas with a higher proportion of non-white population tend to have higher unemployment rates and a lower proportion of employed workers[31].
- Areas with higher unemployment rates may have higher proportions of senior citizens if there is low economic mobility and younger generations move out of the area[32].
- Areas with higher unemployment and lower employment tend to have more people working in construction, service, delivery or production or are rust belt areas that have a declining manufacturing industry[33,34].
- Areas with higher unemployment tend to be poorer and have lower incomes[35] .
- Areas with more senior citizens who are on medicare or those with higher incomes individuals who receive insurance from their employer tend to have lesser uninsured fraction of the population[36,37].
- Counties with higher average incomes tend to have more income inequality and a higher Gini index[38,39].
- Areas with a higher percent of people working in construction, service, delivery, or production may have more poverty[40].
- Counties with more poverty are likely to have less single family homes and more areas zoned for multi-family homes or are likely to contain inner cities and thus have a greater density[41,42]. In such dense areas, there is likely to be shorter commutes and better transit due to these areas being more populous[43,44].

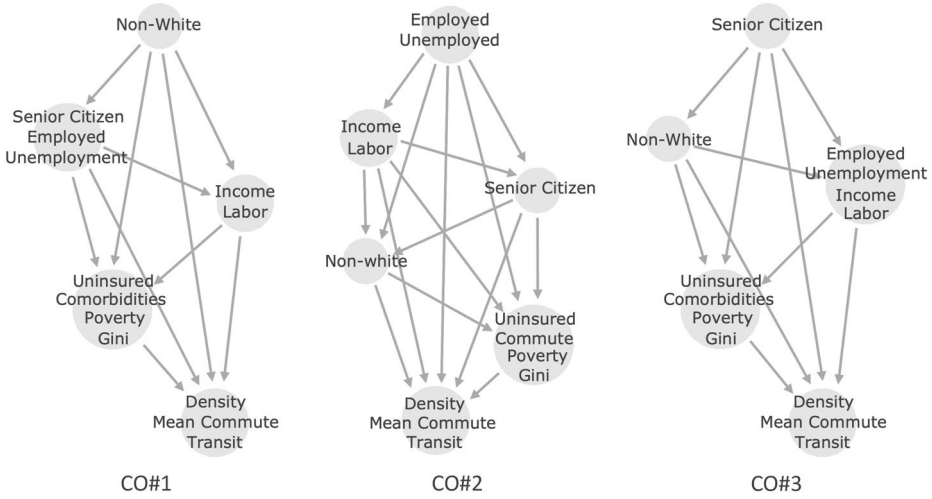

Figure 1. The three primary partial causal ordering that we consider in this work.

CO#2. The second causal ordering is [[Emp, Uemp], [Inc, Lab], SC, NW, [Uins, Com, Pov, GI], [Den, MC, Tran]]. This causal ordering has unemployment, employed, income per capita, and fraction working in manufacturing or manual labor as causing the proportion of senior citizens and non-white fraction of the population. This causal ordering takes into account that senior citizens are more likely to stay in areas with a declining manufacturing industry and little economic mobility[32]. Since senior citizens are less likely to be non-white, a higher fraction of senior citizens causes there to be a lower proportion of non-whites[45].

CO#3. The third causal ordering is [SC, NW, [Emp, Uemp, Inc, Lab], [Uins, Com, Pov, GI], [Den, MC, Tran]]. This causal ordering has the proportion of senior citizens causing the proportion of non-whites. Since senior citizens are less likely to be non-white, a higher fraction of senior citizens causes there to be a lower proportion of non-whites[45]. This causal ordering also has income per capita, the proportion of people in a county working in professions classified as Labor, unemployment, and employment on equal footing. This is the case since manual labor jobs are less temporary and pay lower wages[46].

Evidently, the decision on which causal ordering to choose is quite subjective and requires some prior knowledge of how the various metrics interact with each other. In the case of the problem we are addressing, the causal ordering can even change from region to region depending on the nature of the underlying socioeconomic fabric that governs the social structure. Thus it is important to check how robust the conclusion of our work is vis-à-vis a perturbation in the causal ordering. Ideally, the Shapley Index, S_I (Shapley Index is defined in the "Methods" section), should depend not only on the data but also on the causal ordering. However, S_I should not be agnostic of either the data or the causal ordering since, after all, S_I should encapsulate the information from the causal ordering and the data.

In Fig. 2, we show the result of shuffling the causal ordering CO#1. We randomly rearranging the subgroups into 20 different permutations. The violin plots show the variations in S_I for each metric. The red dot corresponds to the S_I for CO#1. We see that the three most important metric—non-white, density, and mean commute—do not show large variations in S_I. The metrics that are less causally connected to the confirmed case rate, however, show large variation. Nevertheless, there is a clear distinction between the three most important ones and the less important ones. This result shows that while S_I does depend on the causal ordering, the most important metrics are tied to the information that is gleaned from the data. If we restrict ourselves to picking the metric with the highest S_I, and hence, the largest causal connection to the outcome, small perturbations in the causal ordering will not alter our results. This observation displays the robustness of our approach making it not as subjective as it might appear to be on the surface.

Results

The question that we would like to pose is: are there any casual connection between the socioeconomic metrics and the spread of the disease and is this causal connection stable over time and geographies?

We will explore two slices of time in this section. The first one is the period when the disease had just started spreading through the population. In the beginning of 2020, COVID-19 affected the most densely populated parts of the USA, namely, the states in the east coast and some states in the west coast. However other parts of the USA, especially the rural parts of the USA, remained unaffected by large. It was the second wave in mid-2020 that affected the parts that were spared by the first wave with vast majority of the southern states seeing a surge in the number of infected individual. So a study of socioeconomic conditions that might have affected the differential advent of COVID-19 across the social strata should include the months from February 2020 to July 2020. We shall refer to this as Phase I of our study. Consequently, we shall deem Phase II as the next phase of the disease spread from July 2020 to January 2021 keeping it safely before when vaccinations could possibly have

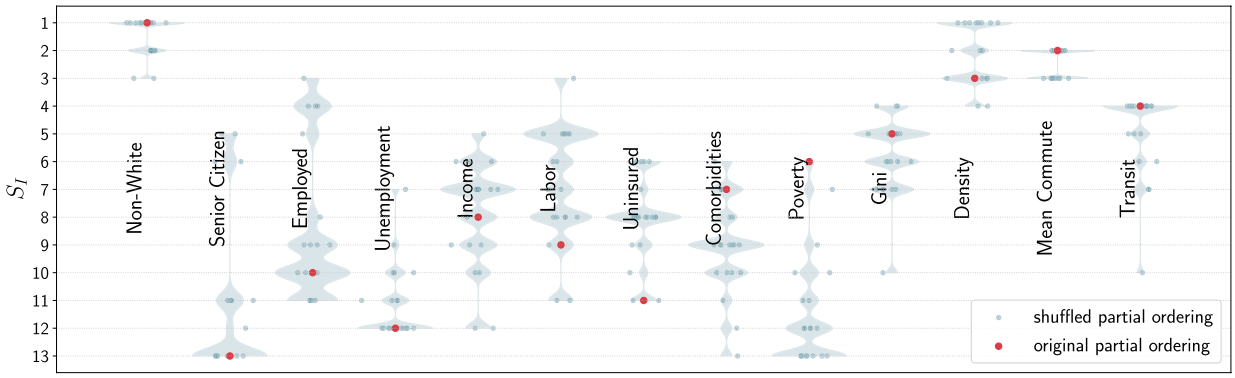

Figure 2. The Shapley Index, S_I, of all the metrics from 20 different causal ordering where the ordering of the subgroups were randomized starting from the causal ordering used in the left panel. The plot shows that the variation in S_I is small for the most important metrics with high $\overline{|S_v|}$ while larger variations in S_I can be seen in the less important metrics depending on the causal ordering. For more detail please see the text.

any effects. We will contrast Phase I against Phase II to understand the role socioeconomic disparities played in the spread of COVID-19.

In Fig. 3 we show the correlations and causation between the socioeconomic factors, population density and the confirmed case rate during Phase I of the disease spread. The network plots in the top panel display the linear correlations between the exogenous and endogenous variable(s). From the correlations in the plots being distinct for each region, one can expect the causal connections between the variables should be different too. This is indeed confirmed by the bar plots, the columns corresponding to the different regions and the rows corresponding to the different causal ordering. A very clear pattern emerges independent of the causal ordering. In the East Coast region during Phase I, the primary causation is driven by the fraction of non-white population, the mean commute and the population density. While the fraction of non-white population has by far the largest causal connection with the confirmed case rates, the mean commute and density variables interchange positions depending on the causal ordering used. However, it is clear that these three are the most important causation.

In the West Coast region, the fraction of non-white and the fraction of senior citizen in a county are the top causation. The fraction employed in a county complete the top three list in the CO#1 and CO#2 while loses importance in CO#3 leading us to believe that its importance is more sensitive to the causal ordering. In the Southern States the conclusion is similar with the fraction of non-white population, the fraction of senior citizen and the population density having the strongest causal connections with the confirmed case rate with non-white fraction being the primary causation.

Moving on to Phase II the pattern is quite different. In Fig. 4 we see that the value of $\overline{|S_v|}$ for all the variables are much lower than the corresponding ones in Phase I for both the East Coast States and the Southern States leading us to conclude that the variables have less effect in causing the spread of the disease which is more homogeneous geographically. The case for the West Coast region is somewhat different with the non-white fraction still holding some causal connection with the spread of the disease along with the fraction of senior citizen for CO#2 and CO#3. In addition, the correlation patterns shown in the network plots in Fig. 4 differ from those in Fig. 3 and seem to be more sparse for the East Coast region and the Southern States while being relatively unchanged for the West Coast region.

Econometric analysis

The machine learning framework along with the causal Shapley values are novel method for analyzing causalities in socioeconomic data. To validate its strengths and to highlight its differences with a more traditional linear causal analysis we use a simple linear model to analyze the effect of socioeconomic metrics on the spread of COVID-19 across the USA.

We use weekly balanced panel data, and we safely assume that standard errors in the balanced panel data are not independent. To begin with, we perform the White test and Breusch-Pagan test to detect the presence of heteroscedasticity. For each of the six datasets (three regions and two phases) we get a p-value < 0.05 from both the tests which leads us to reject the null hypothesis of homoscedasticity and we conclude that heteroscedasticity is present in the residual of this empirical model specification. Furthermore, we perform the Durbin-Watson test for auto-correlation and get test results of < 1 for all datasets leading us to the conclusion that there are significant positive auto-correlations in the datasets. Therefore, we need to use a random effect model with clustered standard error at the county level. These clustered standard errors allow for the presence of heteroscedasticity and correlation in the error term within a cluster. A fixed effect model cannot be used given the socioeconomic metrics do not change over the times during which the COVID-19 prevalence data is collected. The random effect model estimates the effects of time-invariant socioeconomic metrics as presented in the data collected from the US Census Bureau. We build the model with only a subset of the socioeconomic metrics that we consider in the previous analysis since highly correlated metrics cannot be used to fit a linear random effect model. Hence we chose four of the metrics that are not highly correlated and perform the analysis so as to reasonably compare it with the results obtained from the machine learning framework.

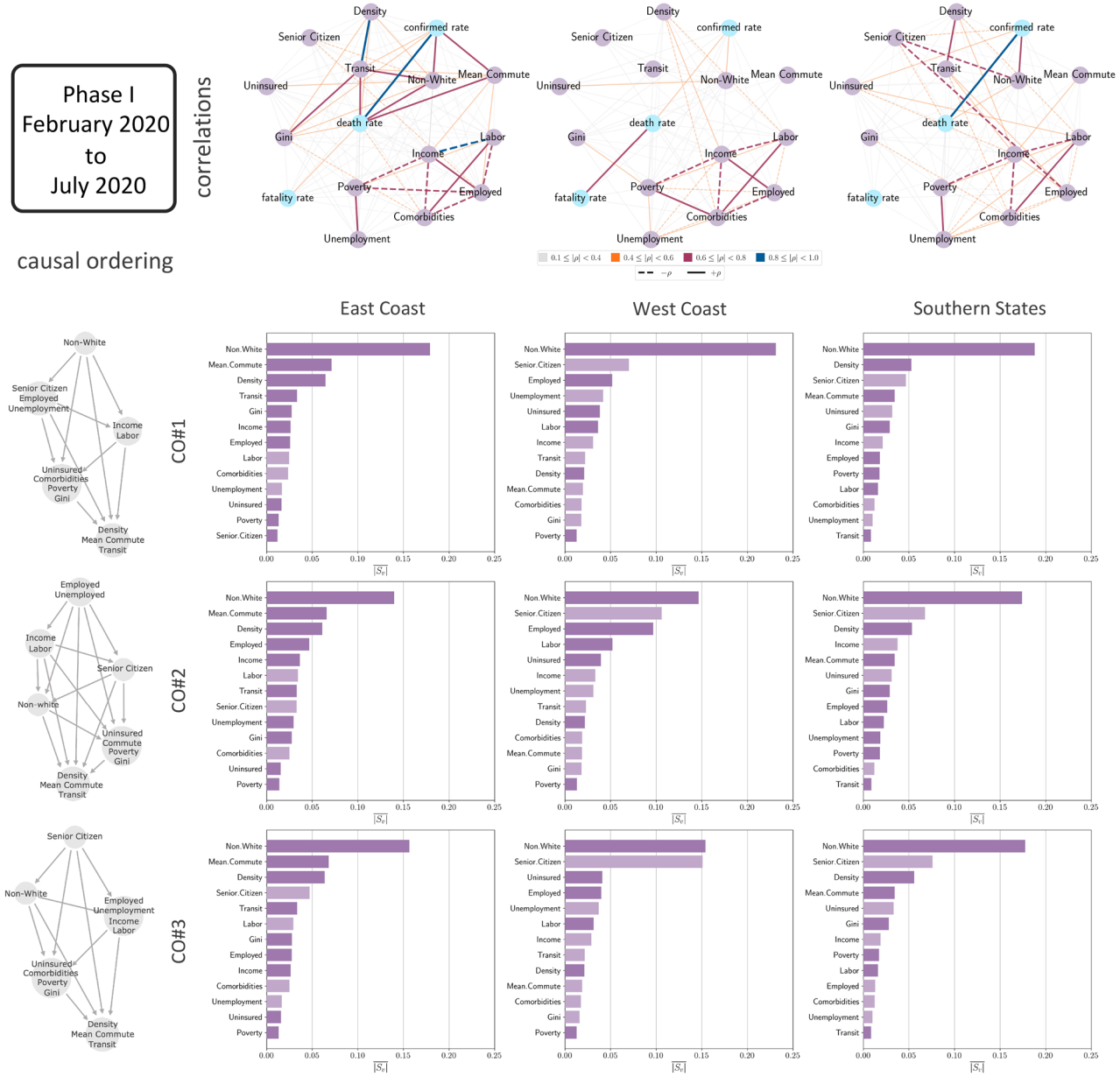

Figure 3. The causal connection between the socioeconomic metrics that we consider in this work and confirmed case rates in various regions of the USA from February 2020 to July 2020. The states included in the East Coast region are District of Columbia, New Jersey, Rhode Island, Massachusetts, Connecticut, Maryland, Delaware and New York. The ones in the West Coast region are California, Oregon and Washington. The states in the Southern States region are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia. The networks in the top row show the correlations between the different metrics in the three regions. The graphs in the left column of the plot show the three partial causal orderings that we consider in this analysis. Each panel of the array of bar plots show the hierarchy of the metrics in terms of mean absolute causal SHAP values, $\overline{|S_v|}$, for a particular region and causal ordering. The lighter shaded bars indicate a negative effect of the metric on the confirmed case rates while the darker bars indicate a positive effect.

The result of the empirical analysis for Phase I is presented in Table 1. The dependent variable is the $\log_{10}$ of weekly COVID-19 confirmed cases per 100,000 individuals within a county. The first row of Table 1 shows that for all the regions, higher population density increases the probability of the spread of the number of COVID-19 cases in the region. The effect is positive and significant (p-value < 0.05). The second row shows that in regions with higher unemployment rate, the spread of COVID-19 is relatively less as seen in the numbers from the west coast (p-value < 0.05). Third and fourth row have the expected results that higher income region and regions with a smaller fraction of non-white population sees a decrease in the probability of the spread of COVID-19 (p-value < 0.05). The effect is significant for all three regions.

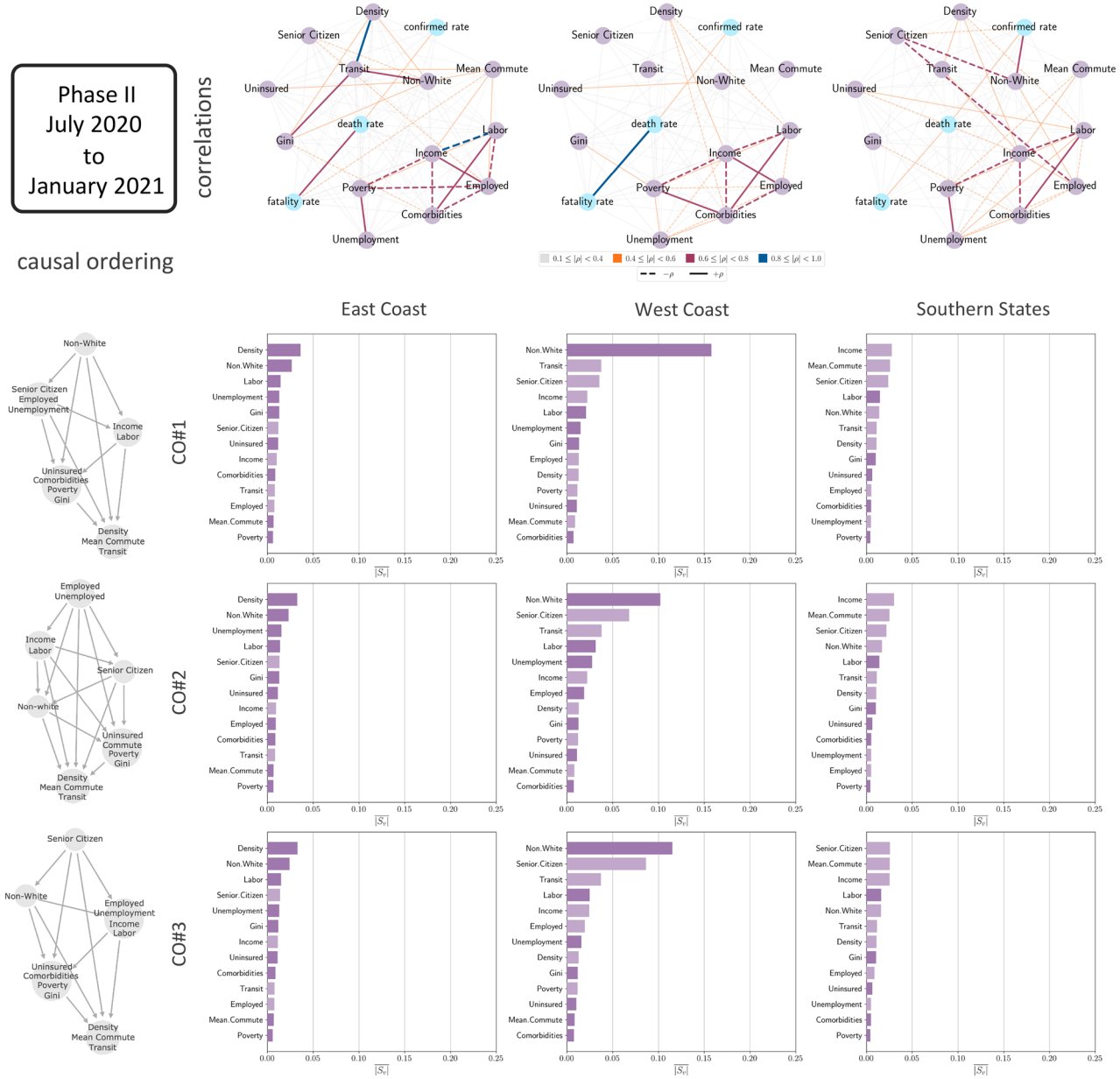

Figure 4. The causal connection between the socioeconomic metrics that we consider in this work and confirmed case rates in various regions of the USA for July 2020 to January 2021. The states included in the East Coast region are District of Columbia, New Jersey, Rhode Island, Massachusetts, Connecticut, Maryland, Delaware and New York. The ones in the West Coast region are California, Oregon and Washington. The states in the Southern States region are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia. The networks in the top row show the correlations between the different metrics in the three regions. The graphs in the left column of the plot show the three partial causal orderings that we consider in this analysis. Each panel of the array of bar plots show the hierarchy of the metrics in terms of mean absolute causal SHAP values, $\overline{|S_v|}$, for a particular region and causal ordering. The lighter shaded bars indicate a negative effect of the metric on the confirmed case rates while the darker bars indicate a positive effect.

The Phase II results in the lower half of Table 1 has similar results as Phase I and shows that socioeconomic metrics affect the spread of COVID-19 significantly, especially in the West Coast states. The primary difference between Phase I and Phase II is that some of the effects of the socioeconomic metrics on the spread of COVID-19 decreases in Phase II compared to Phase I as presented in our previous analysis, especially for the East coast and the Southern States regions. These results highlight two implications of the causal analyses. Firstly, the initial health care policy intervention of some of the US states and the federal government seem to have helped in reducing the differential in the spread of the pandemic from what was seen at the initial stage and was disproportionately affecting the more vulnerable. Secondly, socioeconomic metrics prove to be important in shaping the possible policy interventions to control the spread of the pandemic.

Variables	East Coast	West Coast	Southern States
Phase I			
Density	0.161 (0.037)[1]	0.114 (0.028)[1]	0.170 (0.009)[1]
Unemployment	0.016 (0.024)	−0.094 (0.024)[1]	−0.007 (0.009)
Income	0.036 (0.027)	−0.062 (0.027)[2]	−0.043 (0.010)[1]
Non-White	0.096 (0.034)[1]	0.226 (0.023)[1]	0.160 (0.008)[1]
Constant	1.094 (0.019)[1]	0.675 (0.019)[1]	0.889 (0.007)[1]
Observations	3036	2926	30,668
Phase II			
Density	0.111 (0.046)[2]	0.142 (0.031)[1]	0.1006 (0.006)[1]
Unemployment	0.039 (0.030)	−0.083 (0.027)[1]	−0.0188 (0.006)[1]
Income	−0.035 (0.034)	−0.177 (0.031)[1]	−0.1103 (0.007)[1]
Non-White	0.038 (0.042)	0.226 (0.027)[1]	0.0067 (0.006)
Constant	1.814 (0.024)[1]	1.844 (0.021)[1]	2.1862 (0.006)[1]
Observations	3588	3458	36,224
No. of counties	138	133	1394

Table 1. Results of the causal analyses performed with random effects models. The central value and the robust standard errors (in brackets) are given for the variables that were considered in the analyses.The p-values are marked as: $^1p < 0.01$, $^2p < 0.05$.

Comparison between methods

Having used two very different frameworks for this analysis which point towards a very similar interpretation of the data, we would like to highlight the major differences between the frameworks themselves in an attempt to allow the readers to judge the merits of the novel machine learning framework that we propose for causal analyses.

Non-linearities. Linear models like the random effect model fail to capture non-linearities in the data and their use is further complicated by correlations present amongst the exogenous variables. This leads to the necessity of considering only those variables that are not highly correlated which leaves open the possibility of ignoring a confounding variable. On the other hand, the framework based on ensembles of BDTs that we use naturally models non-linearities in the data taking into account all correlations and, hence, provide a better regression of the data. This allows for the consideration of a larger set of exogenous variables, which in turn, allows for a better modeling of the endogenous variables.

Interpretability. The drawback of using machine learning (beyond linear regression or other simple models) is that they are not easily interpretable and often end up as black-boxes. This is, however, not the case for linear models like the random effects model which allows for a very clear interpretation of the model parameters. The lack of interpretability is, in fact, a major hurdle in the use of machine learning in computational socioeconomics. We address this problem by using Shapley values to interpret the regression model we create with the ensemble of BDTs and do so in a manner in which the casusality in the data is probed, in effect, adding interpretability to a black-box model.

Summary

The work that we present here has two primary components. Firstly, we establish the tenets of a multivariate causal analysis using interpretable machine learning which aims at consistently taking into account non-linearities and correlations and build upon Shapley values, taken from coalition game theory, to extract causal information from the data. Our use of an ensemble of BDTs, instead of other non-linear machine learning models, is motivated by the ease of computation of Shapley values from BDTs using available software packages and should not be taken as a limitation of the causal analysis we propose on a choice of the machine learning models that can possibly be used. The causal Shapley values we calculate are based on the hypotheses about the causal connections between the exogenous variables which are represented by causal chain graphs built on assumptions of the partial causal orderings between the variables. We have tested several hypotheses of these causal orderings to ascertain the robustness of the analysis framework and its dependence on such hypotheses.

Secondly, we use this analysis framework to study the effects of socioeconomic disparities on the spread of COVID-19 in the USA. We use population demographics data gathered from the US Census Bureau and COVID-19 prevalence data from the Johns Hopkins University to fit regression models using the interpretable machine learning framework. The extraction of causal Shapley values from these regression models allow us to infer on the causal connections between the socioeconomic metrics and the prevalence of COVID-19 at the county level. To compare the results of our analysis with a more traditional causal analysis we use a weekly balanced panel

data to fit a random effects model with a reduced set of exogenous variables and find reasonable agreement with the results from the interpretable machine learning analysis.

From these analyses we conclude that:

- The effects of socioeconomic disparities on the spread of COVID-19 was more pronounced at the beginning of the pandemic than at the later stages.
- While in the parts of the USA with higher population density, the spread was driven partially by the population density, socioeconomic metrics like the fraction of non-white population in a county also show significant causal connection with the spread of COVID-19. In fact, population density was not causally connected to the spread of the disease in the West coast region.
- Of particular note is the causal connection between the fraction of senior citizen in a county and the spread of the disease in the West Coast region and the Southern States. The Shapley values being anti-correlated to the COVID-19 confirmed case rate implies that counties with a younger population saw a larger spread of the disease. As we know that the probability of COVID-19 infection increases drastically with age, our results imply that while the older fraction of the population were being differentially affected to a larger extent by the disease it was being spread more widely by the younger and more mobile fraction of the population.

While we have not addressed the question of confounding variables in any detail, it is possible to assume the presence of confounding amongst the variable clustered in a partial ordering. We present this analysis in the Supplementary Information. Assuming the presence of confounding variables does not change the results of our analysis.

With this work, we hope that our proposed methods for causal analysis finds some utility in computational socioeconomics much beyond the application that we have proposed. The conclusions that we draw about the effects of socioeconomic disparities on the spread of COVID-19 in the USA, especially during the onset of the pandemic reinforces what has been observed with clinical data and population level analyses and points to the necessity for restructuring of the crisis response system to nullify the causes of such disparities. We hope our work will lead to more detailed thoughts, insights and actions that will prove to be useful in the near future.

Methods

The data for this work comes from three primary sources:

- The 2019 American Community Survey (ACS) 5-years supplemental update to the 2011 Census found in the USA Census Bureau database for constructing the socioeconomic metrics.
- The population density data that reflects the 2019 estimates of the US Census Bureau.
- Data on COVID-19 prevalence and death rate is obtained from the Johns Hopkins University, Center for Systems Science and Engineering database[47] sourced from github.com/CSSEGISandData/COVID-19.

To study the relevance of socioeconomic condition in the spread of COVID-19, we focus on various metrics that can characterize these at a county level including factors such as per capita income, poverty, the employed fraction of the population, and the unemployment rate. The latter two are not fully correlated since they add up to the fraction of the population that is employable which varies from county to county. We also include factors that relate to the mobility such as the mean commute time in any county along with the fraction of the population that uses a transit system (we do not consider the relative reduction in mobility due to partial lock-downs or closures of institutions). We also include the fraction of senior citizen in a county which can also affect the mobility patterns in the county. For example, a population with a lower median age will, on an average, have higher mobility than a population with a much higher median age. Since studies have already shown that COVID-19 tends to affect the elderly preferentially, the fraction of a population that fall into the senior citizen category (over 65 years of age) is a good measure of quantifying both the mobility and the degree to which age plays a role in determining the spread of COVID-19. To account for professions that put individuals at higher risk of exposure to a larger number of people such as those in service industry, construction, delivery, labor etc., which increases their chance of being infected by COVID-19, we employ a metric that quantifies the fraction of the population that work in these industries per county. Lastly, we include the fraction of individuals in a county that do not have health insurance as a metric to assess if it affects the spread of the disease.

The socioeconomic metrics for each county collected from the US Census Bureau 2019 5-year ACS data used in this work are the following:

- Population density (Den): The population density data taken from 2019.
- Non-White (NW): The fraction of non-white population in any county including Hispanics and Latinos.
- Income (Inc): The income per capita as defined by the US Census Bureau.
- Poverty (Pov): The fraction of the population deemed as being below the poverty line.
- Unemployment (Uemp): The unemployment rate as defined by the US Census Bureau.
- Uninsured (Uins): The fraction of the population that does not have health insurance.
- Employed (Emp): The fraction of the population that is employed.
- Labour (Lab): The fraction of the population working in construction, service, delivery or production.
- Transit (Tran): The fraction of the population who take the public transportation system or carpool excluding those who drive or work from home.
- Mean Commute (MC): The mean commute distance for a person living in a county in minutes.

- Senior Citizen (SC): The fraction of the population that is above 65 years of age.
- Gini index, or Gini coefficient (GI): A measure of the distribution of income across a population used as a gauge of economic inequality.

The abbreviations in parentheses are what we use to refer to these variables while discussing partial causal ordering and the results. A total of nine missing values were found in the data derived from the US Census Bureau. Out of these, eight were for Rio Arriba County, NM and one for Loving County, TX. These were imputed with the corresponding values from the 2011 US Census data. The complete code for data curation can be found in the `Github` repository that is linked in the Data Availability sub-section below. For the data on COVID-19 prevalence we look at the total number of confirmed cases and deaths up until the 15th of January 2021 and define the following rates:

- Confirmed case rate: The total number of confirmed cases per 100,000 individuals in any county.
- Death rate: The total number of deaths per 100,000 individuals in any county.

We do not use the fatality rate (number of deaths per confirmed case) as a measure of the disease spread since in[23] it was shown the fatality rates are mostly uncorrelated with the socioeconomic metrics pointing to the fact that fatality rates were not disproportionately high in certain socioeconomic strata of the society.

Comorbidities are known to aggravate COVID-19 infections leading to higher chances of a symptomatic infection and hospitalization. We include the data provided in a study conducted by the Centers for Disease Control and Prevention[48] which provides the distribution of comorbidities in all the counties in the USA to study the possible effects of comorbidities in determining the extent to which COVID-19 spreads in any region. We define the corresponding metric as:

- Comorbidities (Com): The total fraction of population with one or more pre-existing chronic conditions.

In order to deal with extreme values and outliers, scaling is necessary, especially for an analysis based on linear models and ordinary least squares. While most of the variables collected from the US Census Bureau are expressed in percentages, Labour and Mean Commute commute are in units of US dollars and minutes. None of these variables show large outliers. However, Density shows a large variation over several scales and hence we scale it logarithmically with $\log_{10}$. The same holds true for the confirmed case rate and the death rate. For the econometric analysis using the random effects model, we standardize all the exogenous variables by removing the mean of the variables and scaling to unit variance. For the BDT analysis, we did not need to further scale the data as BDTs are not very sensitive to scaling of data that fall within the same order of magnitude. We explicitly checked this with and without scaled data.

Analysis framework. In this work we aim at extracting the causal connections between the exogenous variables (socioeconomic metrics) and endogenous variables (COVID-19 prevalence and death rate) by using machine learning for modeling underlying multivariate distributions from the data, and then calculating causal Shapley values. The procedure we follow can be delineated as:

- Use an ensemble of boosted decisions trees (BDTs) which act as weak learners that perform reliably in a statistical ensemble. This allows us to build a non-linear model of COVID-19 case and death rates in terms of the socioeconomic metrics and population density at the county level.
- Define the causal flow of the variables by setting up partial causal ordering graphs.
- Calculate the causal Shapley values that allow us to infer upon the causal connections between the exogenous variables and the endogenous ones.

To clarify the analysis procedures we delve into some details of the machine learning framework that we use and the definition of causal Shapley values. Of particular importance is our discussion of how we set up the partial causal ordering graphs that determine the causal flow of the variables. This is the part where an understanding of the interactions between the variables is necessary and we use several hypothesis to make sure we properly address the subjective nature of designing the partial causal orderings.

It should also be noted that we do not use time-series data to train the ensemble of BDT or to extract the Shapley values and, hence, did not require a stationarity test for the dataset. The data used to train the machine learning models is the total confirmed case rates (or death rates) at the end of Phase I and Phase II. The causal Shapley values for each phase are computed from the ensemble of weak learners exploiting the additive property of Shapley values.

Machine learning framework. To perform a regression of the data we use an ensemble of boosted decisions trees (BDTs) taking either disease prevalence or death rate as the endogenous variable while keeping the socioeconomic metrics and population density as exogenous variables. This allows us to not assume a functional form for the model but rely on the data alone. We emphasize here that we are not trying to build a predictive model using machine learning but rather to perform regression in a model-agnostic manner. We use XGBoost[49], a scalable end-to-end boosting system for decision trees that is particularly suitable for sparse data. For data augmentation we use an ensemble of BDTs as weak learners trained on random selections of the sample with

replacement split 70/30 into training and testing sets. This also gives us a stable measure of the accuracy with which we can model the data. We use the coefficient of determination, R^2, calculated from the test sample set aside for each BDT, as our regression accuracy measure and estimate the error of estimating R^2 from the ensemble. It is normal for weak learners to differ in their predictions (it is a consequence of bagging within the ensemble). The prediction of an ensemble of weak learners is taken as the average in a regression problem. The accuracy of the ensemble is represented by the mean R^2.

Causal Shapley values. The Shapley value, first introduced in[24], is a solution concept from cooperative game theory for transferable utility games. The game is characterised by a pair (C, F), where $F = \{1, \ldots, d\}$ is a set representing players in the game, and the *characteristic function* $C : 2^F \to \mathbb{R}$ assigns a non-negative real value $C(S)$ to every coalition $S \subseteq F$, and zero to the empty coalition, i.e. $C(\varnothing) = 0$. The Shapley decomposition is provably (cf.[50], Thm. 2) the only solution concept satisfying the four favorable axioms of *efficiency*, *additivity*, *symmetry* and *null player*, well exposed in[51]. This uniqueness has contributed to the popularity of the Shapley value in the literature of explainable artificial intelligence (XAI). However, different incarnations of Shapley value based explanation methods make different assumptions and use of different approximations, rendering Shapley value based explanations less "unique".

Using Shapley values in a machine learning context amounts to interpreting the players F as features in a model, and the characteristic function C as either the model itself or an evaluation measure of the model's performance. Lundberg et al.[25] identified that several existing explanation methods, including Shapley value based ones, belong to a class of "additive feature attribution methods", and unified Shapley values with the solution concept of local interpretable model-agnostic explanations (LIME)[52], introducing the SHAP model with simplified inputs, i.e. inclusion/exclusion of features. As machine models assume an input shape defined during training, evaluating it in the absence of a feature is not possible in general. Hence, the SHAP characteristic function approximates the (counterfactual) model prediction under removal of features via the expected value of the prediction conditional upon the values of the included features.

In Ref.[25] iteration of so-called Kernel SHAP, this conditional expectation is approximated via the marginal distribution, see Eq. (11) in[25], which amounts to an assumption of feature independence. This is done for computational efficiency, and not inherent to the SHAP framework. The Python SHAP package[53], first released along with[25], is maintained. In[26], Aas et al. suggest improving Kernel SHAP by obtaining the values of the out of coalition features by conditioning upon the excluded features when calculating the expected values. As the conditional distribution is in general not known, Ref.[26] presents four ways of approximating it. One of these is via the empirical distribution of the data, weighing data instances based on a scaled Mahalanobis distance[54]. The authors of Ref.[26] refer to this as conditional Kernel SHAP, which they implement in the R-package shapr[55].

In Ref.[56], Janzing et al. argue that unconditional, rather than conditional, expectations provide the right notion of excluding features, which contradicts the theoretical justification of[25], used in the SHAP software packages. Drawing from Pearl's seminal work on causality[57] (the interested reader is referred to[58] for an introduction to the *do*-calculus), Janzing et al. stress that the *interventional* distribution is represented by the marginal distribution. Leaning heavily on the work of Datta et al.[59], Janzing et al. thus conclude that the marginal distribution—not the conditional distribution—should be used for excluded features in the SHAP calculation.

Shortly after, Heskes et al.[27] show that this marginal, respectively interventional, SHAP calculation only represents direct effects, meaning that 'root causes' with strong indirect effects are ignored in the feature attribution. They introduce *causal* Shapley values by explicitly including the causal relationships between the data, in the SHAP value calculation. They provide an extension to the shapr package, where the user specifies the causal ordering as well as possible confounding, available at[60]. This patch has not yet been included in the official shapr package.

In our work, we will use the form of causal Shapley values defined by Heskes et al.[27] through the definition of partial causal ordering graphs. On one hand, this makes the analysis of the causal connection somewhat subjective and based on our decision of the causal connections between the variables (and the presence of confounding variables). On the other hand, this enforcement of the causal ordering can be seen as a "prior" with which we inform the computation of the causal Shapley values the information that might not be directly gleaned from the data. We will, in particular, use two quantities to examine the data. The first one is the mean of the absolute Shapley values, $\overline{|S_v|}$, that serves as a measure of global importance of a variable[61]. Higher the value of $\overline{|S_v|}$ of a variable, the greater effect the variable has in determining the outcome. We shall also define S_I as the integer index corresponding to the position in a hierarchy that a certain variable holds when ordered by the corresponding $\overline{|S_v|}$ with $S_I \in [1, n]$ for a n-variable problem.

Data availability

All the data used in this work is sources from public databases that have been cited in the main text. All codes that have been used to process and analyze the data can be found at https://github.com/talismanbrandi/Causal-Inference-IML-C19.

Received: 2 February 2022; Accepted: 18 August 2022
Published online: 22 September 2022

References

1. Millett, G. A. *et al.* Assessing differential impacts of COVID-19 on black communities. *Ann. Epidemiol.* **47**, 37–44. https://doi.org/10.1016/j.annepidem.2020.05.003 (2020).
2. Yancy, C. W. COVID-19 and African Americans. *JAMA* **323**, 1891–1892. https://doi.org/10.1001/jama.2020.6548 (2020).

3. Chastain, D. B. *et al.* Racial disproportionality in Covid clinical trials. *N. Engl. J. Med.* **383**, e59. https://doi.org/10.1056/NEJMp2021971 (2020).
4. Moore J.T., Brusca, R. C., & Ricaldi, J.N. *et al.* Disparities in incidence of COVID-19 among underrepresented racial/ethnic groups in counties identified as hotspots during June 5–18, 2020—22 States, February–June 2020. *MMWR Morb Mortal Wkly Rep.*https://doi.org/10.15585/mmwr.mm6933e1 *(2020).*
5. Martinez, D. A. *et al.* SARS-CoV-2 positivity rate for Latinos in the Baltimore-Washington, DC Region. *JAMA* **324**, 392–395. https://doi.org/10.1001/jama.2020.11374 (2020).
6. DiMaggio, C., Klein, M., Berry, C. & Frangos, S. Black/African American communities are at highest risk of COVID-19: Spatial modeling of New York City ZIP Code-level testing results. *Ann. Epidemiol.* **51**, 7–13. https://doi.org/10.1016/j.annepidem.2020.08.012 (2020).
7. Khanijahani, A. Racial, ethnic, and socioeconomic disparities in confirmed COVID-19 cases and deaths in the United States: A county-level analysis as of November 2020. *Ethnic. Health* **26**, 22–35. https://doi.org/10.1080/13557858.2020.1853067 (2021) (**PMID: 33334160**).
8. Pareek, M. *et al.* Ethnicity and COVID-19: An urgent public health research priority. *Lancet* **395**, 1421–1422. https://doi.org/10.1016/S0140-6736(20)30922-3 (2020).
9. Laurencin, C. T. & McClinton, A. The COVID-19 pandemic: A call to action to identify and address racial and ethnic disparities. *J. Racial Ethnic Health Disparities* **7**, 398–402. https://doi.org/10.1007/s40615-020-00756-0 (2020).
10. Goyal, M. K. *et al.* Racial and/or ethnic and socioeconomic disparities of SARS-CoV-2 infection among children. *Pediatrics*. https://doi.org/10.1542/peds.2020-009951 (2020).
11. Wright, A. L., Sonin, K., Driscoll, J. & Wilson, J. Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. in *University of Chicago, Becker Friedman Institute for Economics Working Paper*. https://doi.org/10.2139/ssrn.3573637 (2020).
12. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl. Acad. Sci.* **117**, 19658–19660. https://doi.org/10.1073/pnas.2009412117 (2020).
13. Qiu, Y., Chen, X. & Shi, W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J. Popul. Econ.* **33**, 1127–1172. https://doi.org/10.1007/s00148-020-00778-2 (2020).
14. Stojkoski, V., Utkovski, Z., Jolakoski, P., Tevdovski, D. & Kocarev, L. The socio-economic determinants of the coronavirus disease (COVID-19) pandemic. *arXiv e-prints*arXiv:2004.07947 *(2020).*
15. Ransome, Y., Kawachi, I., Braunstein, S. & Nash, D. Structural inequalities drive late HIV diagnosis: The role of black racial concentration, income inequality, socioeconomic deprivation, and HIV testing. *Health Place* **42**, 148–158. https://doi.org/10.1016/j.healthplace.2016.09.004 (2016).
16. Farmer, P. Social inequalities and emerging infectious diseases. *Emerg. Infect. Dis.* **2**, 259–269. https://doi.org/10.3201/eid0204.960402 (1996).
17. Hosseini, P., Sokolow, S. H., Vandegrift, K. J., Kilpatrick, A. M. & Daszak, P. Predictive power of air travel and socio-economic data for early pandemic spread. *PLOS ONE* **5**, 1–8. https://doi.org/10.1371/journal.pone.0012763 (2010).
18. Quinn, S. C. & Kumar, S. Health inequalities and infectious disease epidemics: A challenge for global health security. *Biosecur. Bioterror. Biodefense Strategy Pract. Sci.* **12**, 263–273. https://doi.org/10.1089/bsp.2014.0032 (2014).
19. Smedley, B. D., Stith, A. Y. & Nelson, A. R. (eds.) *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. https://doi.org/10.17226/12875. (The National Academies Press, 2003).
20. The Council on Ethical and Judicial Affairs. American Medical Association. *Black–White disparities in health care*. JAMA **263**, 2344–2346. https://doi.org/10.1001/jama.1990.03440170066038 (1990).
21. Andrews, R. & Elixhauser, A. Use of major therapeutic procedures: Are Hispanics treated differently than non-Hispanic Whites. *Ethnic. Dis.* **10**, 384–394 (2000). http://europepmc.org/abstract/MED/11110355.
22. Harris, D., Andrews, R. & Elixhauser, A. Racial and gender differences in use of procedures for black and white hospitalized adults. *Ethnic. Dis.* **7**, 91–105. http://europepmc.org/abstract/MED/9386949 (1997).
23. Paul, A., Englert, P. & Varga, M. Socio-economic disparities and COVID-19 in the USA. *J. Phys. Complex.*https://doi.org/10.1088/2632-072X/ac0fc7 *(2021).*
24. Shapley, L. S. A value for *n*-person games. in *Contributions to the Theory of Games* (Kuhn, H. W. & Tucker, A. W. eds.) (Princeton University Press, 1953).
25. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in (Guyon, I. *et al.* Eds.) *Advances in Neural Information Processing Systems* . Vol. 30. 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf (Curran Associates, Inc., 2017).
26. Aas, K., Jullum, M. & Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artif. Intell.* **298**, 103502. https://doi.org/10.1016/j.artint.2021.103502 (2021).
27. Heskes, T., Sijben, E., Bucur, I. G. & Claassen, T. Causal Shapley values: Exploiting causal knowledge to explain individual predictions of complex models. in *Advances in Neural Information Processing Systems* (Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H. eds.). Vol. 33. 4778–4789. https://proceedings.neurips.cc/paper/2020/file/32e54441e6382a7fbacbbbaf3c450059-Paper.pdf. (Curran Associates, Inc., 2020).
28. Minorics, L. *et al.* Testing Granger Non-Causality in Panels with Cross-Sectional Dependencies. *arXiv e-prints*arXiv:2202.11612 *(2022).*
29. Mastakouri, A. & Schölkopf, B. Causal analysis of Covid-19 spread in Germany. in *Advances in Neural Information Processing Systems* (Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. & Lin, H. eds.). Vol. 33. 3153–3163. https://proceedings.neurips.cc/paper/2020/file/205e73579f21c2ed134dbd6ce7e4a1ea-Paper.pdf (Curran Associates, Inc., 2020).
30. Steiger, E., Mussgnug, T. & Kroll, L. E. Causal graph analysis of COVID-19 observational data in German districts reveals effects of determining factors on reported case numbers. *PLOS ONE* **16**, 1–22. https://doi.org/10.1371/journal.pone.0237277 (2021).
31. Immergluck, D. *Neighborhood Jobs, Race, and Skills*. https://doi.org/10.4324/9781351045957 (Routledge, 2018).
32. Mills, B. & Hazarika, G. The migration of young adults from non-metropolitan counties. *Am. J. Agric. Econ.* **83**, 329–340. https://doi.org/10.1111/0002-9092.00159 (2001).
33. Charles, K. K., Hurst, E. & Schwartz, M. The transformation of manufacturing and the decline in US employment. *NBER Macroecon. Annu.* **33**, 307–372. https://doi.org/10.1086/700896 (2019).
34. Lambert, T. E., Mattson, G. A. & Dorriere, K. The impact of growth and innovation clusters on unemployment in US metro regions. *Region. Sci. Policy Pract.* **9**, 25–37. https://doi.org/10.1111/rsp3.12087 (2017).
35. DeFina, R. H. The impacts of unemployment on alternative poverty rates. *Rev. Income Wealth* **50**, 69–85. https://doi.org/10.1111/j.0034-6586.2004.00112.x (2004).
36. Finkelstein, A. & McKnight, R. What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending. *J. Public Econ.* **92**, 1644–1668. https://doi.org/10.1016/j.jpubeco.2007.10.005 (2008).
37. Cunningham, P. J. & Ginsburg, P. B. What accounts for differences in uninsurance rates across communities?. *INQUIRY J. Health Care Organ. Provis. Financ.* **38**, 6–21. https://doi.org/10.5034/inquiryjrnl_38.1.6 (2001).
38. Levernier, W., Partridge, M. D. & Rickman, D. S. Differences in metropolitan and nonmetropolitan U.S. family income inequality: A cross-county comparison. *J. Urban Econ.* **44**, 272–290. https://doi.org/10.1006/juec.1997.2070 (1998).

39. Nielsen, F. & Alderson, A. S. The Kuznets curve and the great U-turn: Income inequality in U.S. counties, 1970 to 1990. *Am. Sociol. Rev.* **62**, 12. https://doi.org/10.2307/2657450 (1997).
40. Adelman, R. M. & Jaret, C. Poverty, race, and US metropolitan social and economic structure. *J. Urban Affairs* **21**, 35–56. https://doi.org/10.1111/0735-2166.00002 (1999).
41. Rothwell, J. T. & Massey, D. S. Density zoning and class segregation in U.S. metropolitan areas. *Soc. Sci. Q.* **91**, 1123–1143. https://doi.org/10.1111/j.1540-6237.2010.00724.x (2010).
42. Kasarda, J. D. Inner-city concentrated poverty and neighborhood distress: 1970 to 1990. *Housing Policy Debate* **4**, 253–302. https://doi.org/10.1080/10511482.1993.9521135 (1993).
43. Bertaud, A. & Richardson, H. W. *Transit and density: Atlanta, the United States and Western Europe*. in *Urban Sprawl in Western Europe and the United Sates* (Urban Planning and Environment, 2004).
44. Levinson, D. M. & Kumar, A. Density and the journey to work. *Growth Change* **28**, 147–172. https://doi.org/10.1111/j.1468-2257.1997.tb00768.x (1997).
45. Bélanger, A. diversity explosion: How new racial demographics are remaking America. *Can. Stud. Popul.***43**, 166. 10.25336/p69s3v (2016).
46. Ono, Y. & Sullivan, D. Manufacturing Plants' Use of Temporary Workers: An Analysis Using Census Microdata. *Ind. Relat. J. Econ. Soc.* **52**, 419–443. https://doi.org/10.1111/irel.12018 (2013).
47. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).
48. Razzaghi H, L. H., Wang Y *et al.* Estimated county-level prevalence of selected underlying medical conditions associated with increased risk for severe COVID-19 illness—United States, 2018. *MMWR Morb. Mortal Wkly. Rep.* **69**, 945–950. https://doi.org/10.15585/mmwr.mm6929a1 (2020).
49. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. 785–794. https://doi.org/10.1145/2939672.2939785 (Association for Computing Machinery, 2016).
50. Young, H. P. Monotonic solutions of cooperative games. *Int. J. Game Theory* **14**, 65–72. https://doi.org/10.1007/BF01769885 (1985).
51. Huettner, F. & Sunder, M. Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values. *Electron. J. Stat.* **6**, 1239–1250. https://doi.org/10.1214/12-EJS710 (2012).
52. Ribeiro, M. T., Singh, S. & Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1135–1144 (2016).
53. Lundberg, S. *Shap 0.39.0*. https://pypi.org/project/ (2021).
54. Mahalanobis, P. C. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. (Calcutta)* **2**, 49–55 (1936).
55. Aas, K., Jullum, M. & Løland, A. *Shapr 0.2.0.9000*. https://rdrr.io/github/NorskRegnesentral/shapr/ (2021).
56. Janzing, D., Minorics, L. & Bloebaum, P. Feature relevance quantification in explainable AI: A causal problem. in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. Vol. 108. *Proceedings of Machine Learning Research* (Chiappa, S. & Calandra, R. Eds.). 2907–2916 (PMLR, 2020).
57. Pearl, J. Causal diagrams for empirical research. *Biometrika* **82**, 669–688 (1995).
58. Pearl, J. The do-calculus revisited. in *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, UAI'12. 3–11 (AUAI Press, 2012).
59. Datta, A., Sen, S. & Zick, Y. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. in *2016 IEEE Symposium on Security and Privacy (SP)*. 598–617. https://doi.org/10.1109/SP.2016.42 (2016).
60. Bucur, I. G. *Shapr*. https://gitlab.science.ru.nl/gbucur/shapr/ (2020).
61. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67. https://doi.org/10.1038/s42256-019-0138-9 (2020).

Acknowledgements

A.P. is funded in part by Volkswagen Foundation within the initiative "Corona Crisis and Beyond—Perspectives for Science, Scholarship and Society", Grant number 99091. I. S. is grateful for support received through the EXAIGON project from industrial partners and the Research Council of Norway (Grant no. 304843). This research was supported in part through the Maxwell computational resources operated at DESY, Hamburg, Germany.

Author contributions

All authors contributed in their own capacities in performing the analysis and writing the draft. The contribution of V.S. deserves a special mention. He contributed significantly to the conceptual design and implementation of the causal analysis, the building of the causal orderings and data analysis.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-18725-4.

Correspondence and requests for materials should be addressed to A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.