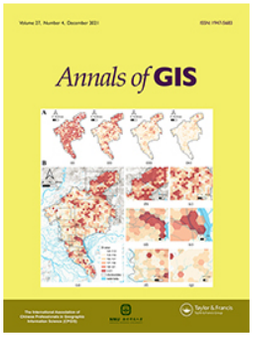

Annals of GIS

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tagi20

COVID-19 impact on excess deaths of various causes in the United States

Akhil Kumar, Yogya Kalra, Weihe Wendy Guan, Vansh Tibrewal, Rupali Batta & Andrew Chen

To cite this article: Akhil Kumar, Yogya Kalra, Weihe Wendy Guan, Vansh Tibrewal, Rupali Batta & Andrew Chen (2021): COVID-19 impact on excess deaths of various causes in the United States, Annals of GIS, DOI: 10.1080/19475683.2021.1982001

To link to this article: https://doi.org/10.1080/19475683.2021.1982001

Published online: 29 Sep 2021.

ANNALS OF GIS
https://doi.org/10.1080/19475683.2021.1982001

Taylor & Francis
Taylor & Francis Group

COVID-19 impact on excess deaths of various causes in the United States

Akhil Kumar[a], Yogya Kalra[b], Weihe Wendy Guan[c], Vansh Tibrewal[d], Rupali Batta[e] and Andrew Chen[f]

[a]Turner Fenton Secondary School, Canada; [b]Central Peel Secondary School, Canada; [c]Center for Geographic Analysis, Harvard University, United States; [d]Dhirubhai Ambani International School, India; [e]Notre Dame High School, Lawrenceville, San Jose, United States; [f]Burnaby South Secondary School, Canada

ABSTRACT

Media regarding COVID-19 fatality counts is crucial, affecting policy and health measures nationwide. However, misinformation regarding other causes of death has led to dubious claims about the seriousness of the coronavirus. This research aims to identify the changes in a dozen causes of death during the pandemic using CDC data from 1999 to 2020. Using the Exponential Triple Smoothing (ETS) algorithm, this project estimated the mortality of eleven causes of death for 2020 under the assumption of no COVID-19 pandemic. Using Power BI and Tableau, this data was visualized together with 2020 actual death counts to determine which causes of death were significantly impacted by the coronavirus. The dashboard revealed an increase in several causes of death including Alzheimer's Disease and Diabetes, a decrease in Chronic Lower Respiratory Disease deaths, and a slight increase in Influenza deaths. These findings, while at odds with much of the media surrounding COVID-19 mortality, are corroborated by adjacent scientific research.

ARTICLE HISTORY
Received 11 April 2021
Accepted 12 September 2021

KEYWORDS
COVID-19; spatiotemporal; excess deaths; misinformation; fatality; pandemic

Introduction

COVID-19 has disrupted lives around the world, making changes everywhere from public health to politics. In the United States, the fatality count for COVID-19 is essential in policy discussions, influencing state-wide responses to shelter in place orders and vaccine distribution. From Congress to the public eye, mortality rates due to COVID-19 have been closely scrutinized in their highs and lows, though most of the attention has been focused on patients who died with a confirmed COVID-19 infection. However, it is suspected that the true impact of COVID-19 on mortality goes far beyond the virus itself. The pandemic may have impacts on mortality across various causes. In spite of the vast media coverage of the pandemic, there is little direct reporting on the extent of such impacts. Using CDC data from 1999 to 2020, this paper models and analyzes the excess deaths under various causes during the COVID-19 pandemic.

The problem – misinformation

There has been a large variety of misinformation surrounding COVID-19 shared online through the pandemic – from its origins to its danger (Schaeffer 2020). Such misinformation amplifies fears of preventative measures, technology and data (Enders et al. 2020). A particular target of misinformation is the accuracy and interpretation of death counts released by the CDC, with many people believing that the numbers have been vastly overestimated by the CDC (Brown 2020). This controversy stemmed from CDC's data revealing that only 6% of all 'COVID-19' categorized deaths listed the virus as the sole cause of death on the victim's death certificate. This led to accusations that health authorities were classifying those that 'test positive' for the virus 'but die of another pathology' as COVID-19 deaths (Susman 2020).

The 6% datapoint meant that 94% of COVID-19 death patients had died because different conditions had compounded with COVID-19 to become fatal. This is quite normal, as over two-thirds of all deaths had multiple causes of death in 2015 (Koch 2015). The 6% of COVID-19 deaths with only one cause of death listed (COVID-19) may have been a result of a rare occurrence, or even an improper death certificate. A study conducted looking into accuracy of death certificates found that 48% had one of five errors in the 'cause of death' section – unacceptable cause errors, non-specific errors, incorrectly completed errors, irrelevant information errors, and incorrect order errors (Cambridge and Cina, 2010). The other cause for this may be rare occurrences where very young individuals with no pre-existing conditions die from COVID-19. Misleading information and disinformation over the determination of both excess and direct

CONTACT Rupali Batta rupalibatta04@gmail.com Notre Dame High School, 1962 Wayne Cir, San Jose, CA, 95131

COVID-19 deaths have resulted in a public misunderstanding over both the nature and meaning of publicly reported numbers.

Current literature on the overestimation of COVID-19 deaths is limited to discussion surrounding potential estimation biases. Previous research identified sampling biases that led to lawmakers from the Congressional House Oversight and Reform committees being incorrectly informed that COVID-19 had 10 times higher mortality than seasonal influenza in a March 11 Congressional hearing (Brown 2020). The researchers found that confusions between infection fatality rates and case fatality rates led to the miscalculation of predicted COVID-19 mortality. The confusion had stemmed from a New England Journal of Medicine editorial that miscited the infection fatality rate of seasonal influenza, as 0.1%, which was then compared to an adjusted COVID-19 mortality rate of 1% (Fauci, Lane, and Redfield 2020). Brown hypothesized that the mentioned sampling bias likely resulted in harsher mitigation tactics to combat the spread of the virus, which in turn created further psychological harm and economic disruption to the population. Similarly, researchers at Harvard University explored the prevalence of COVID-19 misinformation on overestimated mortality rates, finding that nearly 30% of Americans support the theory that coronavirus-related deaths have been exaggerated (Enders et al. 2020).

There is a glaring lack of literature aimed at identifying and disproving the many myths, conspiracy theories and general disinformation shared and distributed digitally. The influence of online misinformation on the beliefs of the general public is profound. According to a survey from the Pew Research Centre, 48% of Americans said they have viewed at least one article of completely false news. Of that news, respondents identified roughly 41% of the stories as regarding the magnitude of risks in the pandemic, with 22% of the stories claiming that the risk was high and 15% claiming that the risk was low. In line with these results, a whopping 62% of respondents believed that the risk of COVID-19 as portrayed in the media was exaggerated (Pew Research Center, 2020). One of the most prominent ways this misinformation spreads is through social media. A study conducted to measure the rate of misinformation on Twitter found that approximately 3 million people spread false news, and that misinformation spreads six times faster than truth (Vosoughi, Roy, and Aral 2018). This illustrates the alarming rate at which myths, conspiracy theories and falsehoods spread on social media.

Another example of misinformation comes from a now-retracted John Hopkins student-written article similarly describing 'the number of deaths by COVID-19' as 'not alarming,' with 'relatively little effect' of COVID-19 on other causes of death. The article, posted on November 22nd, 2020, quickly spread on social media through COVID-19 sceptics and was later taken down by the newsletter on November 26th, 2020. The article was based on Genevieve Briand's (Assistant Director for Johns Hopkins University's Applied Economics Master's Degree Programme) presentation, titled 'COVID-19 Deaths: A Look at U.S. Data.' Briand was quoted saying 'All of this points to no evidence that COVID-19 created any excess deaths. Total death numbers are not above normal death numbers'. She also claimed that deaths from other causes have been miscategorized as COVID-19 deaths. These accusations were quickly shut down, but their impact had propagated through the society (Oransky 2020).

Research objective

Through data exploration, analysis and visualization, this study aims to provide a better understanding of the excess deaths during the COVID-19 pandemic-pledged year of 2020. Excess deaths are defined as the difference between expected deaths if there were no COVID-19 pandemic and observed deaths. Previous studies of excess deaths found that publicly reported COVID-19 deaths may be an underestimate – failing to account for non-COVID-19 deaths caused by the consequences of the pandemic but not directly by the virus infection. These deaths were classified by the biological marker/root cause that led to the death (i.e. Alzheimer's, Cancer, Heart Disease) but lacks the circumstance that led to their death (Centers for Disease Control and Prevention 2021). Thus, this research determines the relevance of such deaths to the pandemic using space-time co-occurrence with COVID-19 death counts. This study analyzes 2020 mortality data (all causes of death including COVID-19) in correlation with mortality data from 1999 to 2019 to come to its conclusions. Using data visualization software, a series of spatiotemporal dashboards were made, providing mortality trends for each of the 50 states, on a month-to-month basis for the years from 1999 to 2020, to debunk misinformation.

This study is set off to explore if COVID-19 related deaths are drastically underestimated by the public, guided through myths, propaganda, and inaccurate information. Furthermore, this research will identify which specific causes of death have exceeded the expected numbers, which have fallen short of the

expected numbers, and which seem unaffected by COVID-19. By examining such second-order impacts and their constituting deaths, this research may help the public become more aware of the true impacts of the COVID-19 virus.

Hypotheses

Our first hypothesis is that COVID-19 death counts have not been overestimated, despite public sentiments to the contrary, and that the pandemic has led to more deaths among other pathologies too. Throughout the pandemic, the circulation of online misinformation has led many to believe in a vast overestimation of COVID-19 deaths in CDC death counts. This scepticism primarily stems from the alleged misclassification of COVID-19 deaths, of which a large amount is claimed to be from other pathologies. As such, many use these claims to discredit the overall risk of the disease and its effect on excess mortality. Researchers have found that the spread of these beliefs lead to lessened compliance with lockdowns and other preventative measures, further increasing the risk of transmission and hospitalization (Tasnim et al. 2020).

The second hypothesis is that in the earlier months of the pandemic when COVID-19 testing was not widely available, excess deaths categorized under other causes would spike because of misclassification. Such spikes would level off after COVID-19 testing became available. Coronavirus testing facilities were overwhelmed during February, March and April 2020, creating shortages in Los Angeles (Baumgaertner & Karlamangla, 2020), San Francisco (Vaziri 2020), Dallas (Root 2020), Chicago (Lourgos and Heinzmann 2020), and dozens of other major cities. Rural areas had little to no testing whatsoever (Lovett and Frosch, 2020). This limited testing may have resulted in deaths being classified incorrectly as, for example, Alzheimer's Disease instead of COVID-19. A study by Yale University found that many excess death spikes occurred weeks before testing was widely available (Weinberger et al. 2020).

The third hypothesis is that the pandemic caused non-COVID-19 deaths that could have otherwise been avoidable. The pressure around healthcare systems accompanied by patients' fear of the virus may have led those who need medical care (for non-COVID-19 related conditions) to not be able to access it, or not receive adequate care needed. This was especially prevalent when and where COVID-19 death spikes, such as during the beginning of the pandemic in coastal cities like New York, where healthcare systems were overwhelmed. According to the CDC, approximately 41% of U.S. adults had delayed or avoided medical care, with 12% being emergency or urgent care (CDC, 2020).

The fourth hypothesis is that respiratory diseases may have a decrease in deaths due to coronavirus prevention policies. Regulations on avoiding close physical contact, increased vigilance about touching public appliances, mask-wearing, and the shelter-in-place orders all decrease the risk of contracting respiratory diseases. 'Any precaution you take to avoid COVID will also reduce your risk of contracting an influenza virus.' said Dr. Casey Kelley, founder of Case Integrative Health in Chicago (Healthline 2020). A study by the University of Maryland in 2013 found that mask-wearing reduces aerosols shedding of virus (Milton et al. 2013). Furthermore, airborne influenza circulates especially quickly in elementary schools (Coleman and Sigler 2020) and in offices (Blue 2013). The lockdown order limits contamination in those regions.

The last hypothesis is that dips in other causes of death may correlate with spikes in COVID-19 mortality. Underlying conditions such as Diabetes mellitus, chronic lung disease, and cardiovascular disease have a higher risk for severe reactions from COVID-19 (CDC COVID-19 Response Team 2020), making them more susceptible to hospitalizations and death. This increased risk may have resulted in more critically ill patients of other conditions die of COVID-19 death, thus reducing the death count for other pathologies.

This study serves to fill the gap in clarifying the misinformation, through analysis of official data, applying scientific models, and creating visualization dashboards, to empirically support or disprove these hypotheses.

Planned approach

Previous literature revolving around excess deaths was predominantly focused on the beginning of the pandemic – from January to March 2020. Many of these studies looked at New York specifically, where it was found that 22% of all deaths were excess deaths. Weinberger concluded that the COVID-19's pandemic deaths were severely underestimated (Weinberger et al. 2020). The research attributed these underestimated deaths to four main factors: lack of access to diagnostic testing, false negatives from COVID tests, infections after a negative test, and a misdiagnosis by physicians. The study used five years of data to estimate what 2020 deaths should have been for the time, then took the difference between this estimation and actual deaths. Another similar study conducted by Woolf used a Poisson Regression Model to predict and analyse

2020 mortality for the months of March to July (Woolf et al. 2020). The study looked at ten different causes of deaths and concluded that 17% of deaths were excess, with 35% of them being non COVID-19 related.

This type of literature, though providing valuable perspectives and insights into the trends of excess deaths, were limited to the early part of the pandemic, used just a few years of historical data for the 2020 mortality prediction, and had no forms of data visualization. For instance, while Lee and Andris constructed a county-level dashboard of excess deaths, they did not separate those mortalities by cause nor include comprehensive data encompassing a whole year under the pandemic (Lee and Andris 2020). This study aims to fill these gaps by using data from 1999–2019 to predict 2020 mortality without pandemic, creating plentiful visualizations, and analysing excess deaths for the entire calendar year of 2020 at a far more granular level in terms of different underlying causes of death.

Data and methodology

Data source

The data that was used in this study was acquired from the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS), which includes the National Vital Statistics System (NVSS) that has the mortality statistics and access to the CDC WONDER data platform. Within the WONDER system, this study used the underlying causes of death for 1999–2019 by bridged-race categories (Centers for Disease Control and Prevention 2020a). The data for 2020 is provisional and was acquired from data.cdc.gov within the NCHS data platform (National Center for Health Statistics 2021). The linkage between 1999–2019 and 2020 data was done through the ICD-10 codes provided by the CDC. See Table 1 for reference. The natural cause of death was excluded as no ICD-10 code was present and COVID-19 was only present in 2020 data files.

There are other data sources such as the Human Mortality Database (Human Mortality Database 2020) and the United States Mortality Database (United States Mortality Database 2020) but none of these sources provide death counts for specific causes of death like Diabetes or Alzheimer's, thus not used in this study.

The data provided by the CDC for 2020 is the most widely used dataset for analysing excess deaths and mortality in 2020 and now 2021. As of March 21st, 2021, it has 715,000 views and 132,000 downloads (NCHS, 2021). This dataset was used by the New York Times in conducting their excess death study (The New York Times 2020) as well as in Dr. Steven H. Woolf's study (Woolf et al. 2020) and the CDC's own excess death study (National Center for Health Statistics, 2021). The novelty that our study brings is studying the excess deaths in specific causes of death.

Mortality data for 1999–2019 was downloaded from CDC WONDER for each state, for each cause of death for each month whereas 2020 data only came in weekly counts for each state and cause of death. Additionally, this study also downloaded yearly mortality data for each cause of death for each state for 1999–2019 (WONDER, 2020). In addition to mortality data, this study also used population estimates from 1999 to 2019 from CDC WONDER and the postcensal population estimates for 2001–2009 (CDC Wonder 2021a). The 2019 estimated population was applied for 2020 since 2020 population data was not available at the time of this study.

Table 1. Underlying causes of death linking 1999–2019 with 2020 data and our dashboard.

Causes of Death Names in 2020 Dataset	ICD-10 Code in 1999–2020 Dataset	Short Names in our data files	Causes Names on our Dashboards and Tables
All Cause	N/A	ALL	All Cause
Septicaemia	A40-A41	SP	Septicaemia
Malignant neoplasms	C00-C97	MN	Malignant Neoplasms
Diabetes mellitus	E10-E14	DB	Diabetes Mellitus
Alzheimer disease	G30	AD	Alzheimer's Disease
Influenza and pneumonia	J09-J18	IP	Influenza and Pneumonia
Chronic lower respiratory diseases	J40-J47	CLRD	Chronic Lower Respiratory Disease
Other diseases of respiratory system	J00-J06, J30-J39, J67, J70-J98	ORSD	Other Respiratory System Diseases
Nephritis, nephrotic syndrome and nephrosis	N00-N07, N17-N19, N25-N27	MNSN	Nephritis, Nephrotic Syndrome and Nephrosis
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00-R99	UD	Unclassified Death
Diseases of heart	I00-I09, I11, I13, I20-I51	DH	Diseases of the Heart
Cerebrovascular diseases	I60-I69	CD	Cerebrovascular Disease
COVID-19, Multiple Cause of Death[a]	U071	CVDM	COVID-19 MCP
COVID-19, Underlying Cause of Death[a]	U071	CVDU	COVID-19 UCP

[a]Note: COVID-19 was only found in the 2020 data files.

A major limitation of these datasets was the abundance of suppressed values in 1999–2019 and 2020 datasets for other causes of death (not all causes). The explanation of how these suppressed values were treated can be found in the methodology section below.

There must be a differentiation made between underlying causes and the listed cause of death. The CDC defines the cause of death as being 'the disease or injury which initiated the train of events leading directly to death' (CDC, 2019). Therefore, if an Alzheimer's patient contracted and died from COVID-19, their cause of death would be COVID-19. It is the same reason that while diabetes is a specified comorbidity, a patient with diabetes who dies from COVID-19 has the cause of death listed as COVID-19. Similarly, if diabetes caused a patient's death, even though the patient may have tested positive with COVID-19, as long as COVID-19 is not the leading cause of death, the death would be categorized as a death of diabetes, which is the underlying cause of death. Such death would also be counted under COVID-19 Multiple Cause of Death, but it will not be double counted in the total death count.

The cause of death in this study is based on CDC's definition of Underlying Cause of Death, which was selected from the conditions entered by the physician on the cause of death section of the death certificate. When more than one cause or condition is entered by the physician, the underlying cause is determined by the sequence of conditions on the certificate, provisions of the ICD, and associated selection rules and modifications. Each death certificate contains a single Underlying Cause of Death, though it may have up to twenty Multiple Causes of Death (CDC Wonder 2021b).

In this study, only COVID-19 has an additional separated count for Multiple Cause of Death. These are deaths attributed to a different Underlying Cause of Death, where COVID-19 was a contributing factor, though may or may not be ranked as the top factor. These COVID-19 Multiple Causes of Death counts are not included in the total death counts because they are already counted under their respective main underlying causes.

Estimating suppressed values

The first challenge encountered in this study concerned the treatment of suppressed values. The CDC suppresses death counts for any given cause-state-month (or week) when the count is below 10, for privacy protection purposes (CDC Wonder 2021a). Estimating these suppressed values is the first step in preparing the historical data for the simulation of 2020 mortality, and for making 2020 actual data comparable to the simulated results.

Table 2. Distribution of suppressed values among causes of death.

Causes of Death	1999–2019 # Suppressed	2020 # Suppressed	1999–2019% Suppressed	2020% Suppressed
Septicaemia	2557	1233	19.90%	44.74%
Unclassified Death	2489	1209	19.37%	43.87%
Other Respiratory System Diseases	2178	1175	16.95%	42.63%
Nephritis, Nephrotic Syndrome and Nephrosis	1760	1061	13.69%	38.50%
Influenza and Pneumonia	1085	1094	8.44%	39.70%
Alzheimer's Disease	534	452	4.15%	16.40%
Diabetes Mellitus	390	621	3.03%	22.53%
Chronic Lower Respiratory Disease	103	410	0.80%	14.88%
Cerebrovascular Disease	25	385	0.19%	13.97%
Malignant Neoplasms	0	3	0.00%	0.11%
Diseases of the Heart	0	2	0.00%	0.07%
COVID-19 MCP		326		11.83%
COVID-19 UCP		332		12.05%
All Causes	0	0	0.00%	0.00%
Total	**11,121**	**8303**	**7.21%**	**21.52%**

In the 1999–2019 dataset, out of the 154,224 total data points (21 years monthly per state per cause), 11,121 or 7.2% are suppressed. Among the 51 states (including the District of Columbia), 28 states have at least one suppressed value. Of the 11 causes, 9 have suppressed values across the nation, the exceptions are Malignant Neoplasms (Cancerous Tumours) and Diseases of the Heart, as shown in Table 2.

In the 2020 dataset, out of the 38,584 total data points (one year weekly per state per cause), 21.5% are suppressed. All 53 geographic units (50 states, the District of Columbia, Puerto Rico, and New York City) have at least one suppressed value. All 11 causes have suppressed values across the nation, the exception is the 'All causes' category, as shown in Table 2.

Geographically, in both datasets, most suppressions happen in the states with smaller population sizes, as shown in Figure 1.

Past work endorsed by the CDC on this topic explains the impact of suppressed values on local mortality rates using CDC WONDER (Tiwari et al., 2014). Explanations of using two Bayesian models on estimating county-level suppression can be found in Quick (2019). Both of these papers estimate suppressed counts at the county level using state data and do not fit the requirements of this study that is to estimate the monthly mortality rates at the state level for various causes of death.

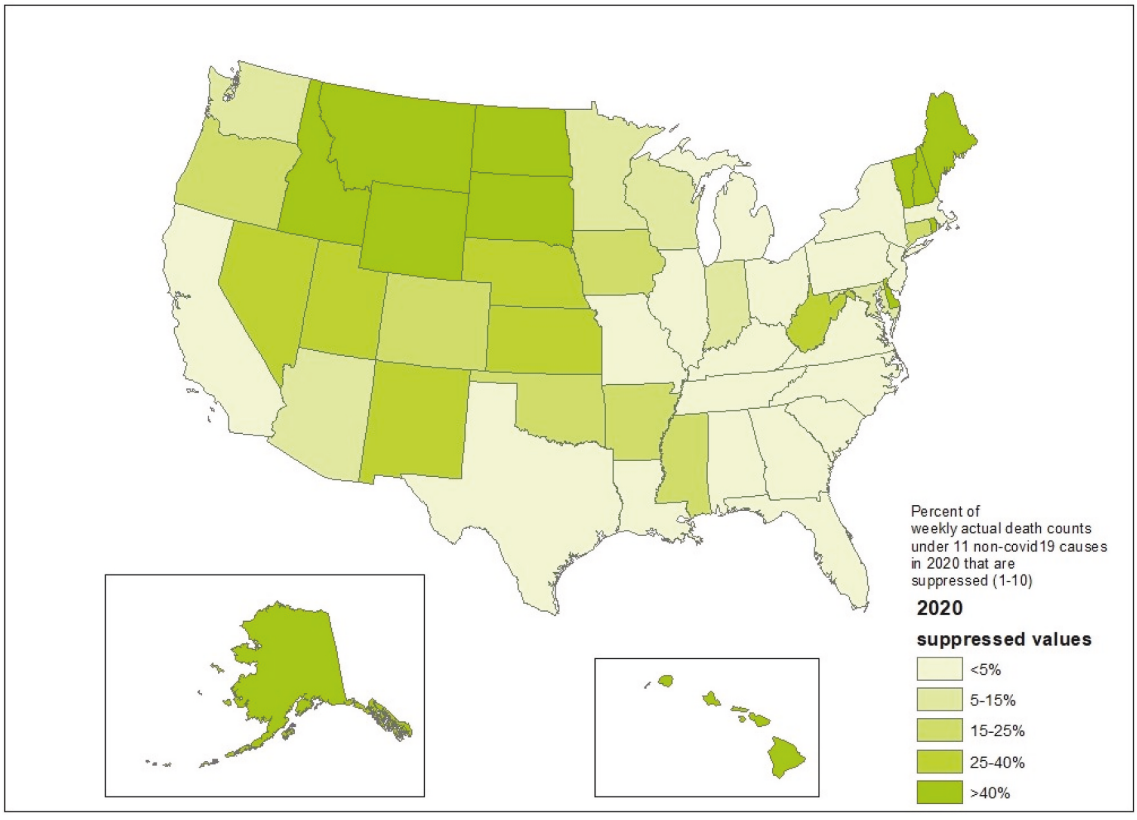

Figure 1. Spatial distribution of suppressed values.

To estimate the suppressed values this study treated each cause of death for each state separately. The first step was to calculate the percent average of the historical data for each month during 1999–2019. This was taken as an average of the 21 years of data for each cause of death for each state for each month. Step two is to find the residual of the suppressed values for each year, which was taken from the difference between the yearly and monthly data for each state and for each cause of death. Step three distributes the residual according to the weight calculated from step one, which is the percent average for each month. If all the values for a particular month are suppressed across the 21 years, this study took the percent average for that month from the nearest neighbouring state as the residual distribution weight. The nearest state is determined by the distance between polygonal centroids of two states, calculated in ArcGIS Pro (Esri 2021). Overall, 2,814 values (25% of the total suppressed values for 1999–2019) were computed using neighbouring state percentages.

For suppressed 2020 death values, the residual values were calculated from the total United States values for each cause of death. The residual was distributed across the suppressed states normalized by 2019 state population.

To compare with monthly data from 1999 to 2019, the 2020 weekly data, after filling in suppressed values by estimations, are divided evenly into daily values, and then aggregated into monthly counts. This process is repeated by cause and by state. The death rate per 100,000 people is then computed using the 2019 state populations.

Computing the prevalence metrics

After estimating the suppressed values, this study used the PHM value estimation methodology developed for 'Geo-visualizing Diet, Anthropometric and Clinical Indicators for Children in India: Enabling District Prioritization for Interventions' (Subramanian et al. 2020) to calculate the prevalence-headcount metric. To account for temporal variations, the values were normalized and calculated for each month individually. The steps to calculate the PHM value were to first normalize the prevalence and headcount metrics then combine them together after the normalization. The three formulas are the following according to the aforementioned paper:

Step 1 – Normalizing Prevalence: $P^{norm} = (P - P_{min})/(P_{max} - P_{min})$

Step 2 – Normalizing Headcount: $H^{norm} = (H - H_{min})/(H_{max} - H_{min})$

Step 3 – Calculate PHM: $PHM = (P^{norm} + H^{norm})/2$

The prevalence and per capita values were computed with population estimates from 1999 to 2019, in which the 2019 population was used for 2020 and postcensal population estimates were used for 2001–2009, taken from CDC WONDER (CDC Wonder 2021a).

The raw count of deaths, when reported by state, cannot directly reveal the pattern of severity of the pandemic. States ranked high with the total death counts are inevitably states with larger populations. The per capita death rate gives a comparable measure of severity between states. However, the impact of the pandemic on densely populated states is different than on sparsely populated ones, even when the per capita death rate is similar. The PHM value provides comprehensive metrics to compare severity of the pandemic's impact, combining total counts and per capita rates.

Predicting 2020 mortality without COVID-19

To predict the 2020 mortality values from 1999 to 2019 data, this study used the Exponential Triple Smoothing (ETS) algorithm which accounted for seasonality in the dataset (Microsoft 2019a). As stated in the Microsoft Power BI Team's technical blog: 'The algorithm (ETS AAA) is a state-space-based forecasting method. Essentially, forecasts are weighted averages of past observations, with recent observations given more weight.' (The Power BI Team, Microsoft 2021)

To ensure the model has adequate accuracy in its prediction, a simulation test was done for 2019 causes of death. The same methodology was applied with the exception of taking out 2019 death data from the input, having the predictive death values for 2019 be the output. This mirrors what was done with the 2020 predictive model.

The results of this 2019 prediction vs. actual comparison are available for public examination at the dashboard 'Mortality of Various Causes in the United States – 2019 Prediction vs. Actual Comparison' (Kumar and Tibrewal 2021). Since 2020 values are under the impact of COVID-19, this 2019 accuracy test was designed to ensure that the model was performing adequately and that the predictions would be reliable.

Compared with the model predicted death values from 1999 to 2018 data, 380 actual values in 2019 (5.6% of the total compared values) fell beyond the 95% confidence range of the predicted values. They are almost evenly split, 2.75% fell below the range, and 2.9% rose above the range. The same model was then used to predict 2020 death values from 1999 to 2019 data. The actual values in 2020 were compared with the predicted values cause-by-cause, state by state, and month by month. Five percent (5%) of the 2020 actual values fell below the 95% confidence interval of the prediction, while 14.5% rose above it. More details are presented in the result section below.

The dashboard for US 2020 mortality prediction-actual comparison is part of the 'COVID-19 Impact on Mortality of Various Causes in the United State' dashboard (Kumar and Tibrewal 2021), which displays the forecasted values per 100,000 population for each cause of death for each state.

An additional feature on the dashboard is to detect anomalies in the 1999–2019 dataset. The anomalies were computed via the SR-CNN algorithm, a built-in algorithm in Power BI by Microsoft and is configured to account for 90% sensitivity, which means data points that do not fit within a 10% change range among the total data points (per cause, per state and per month during the 1999–2019 period) will be marked as an anomaly (Microsoft 2019b). This helps dashboard users to examine irregular changes in the data and their volatility, on top of seasonal fluctuations and long-term trends.

Data processing and visualization

Python scripts and Excel formulas were applied to compute data. All the data processing steps and files can be found on the Harvard CGA GitHub page (Kumar and Tibrewal 2021). The final derived dataset can be found on Harvard Dataverse (Center for Geographic Analysis 2021).

Due to the complexity of the data, several dashboards were built to facilitate data visualization and exploration for the research team. At the same time these dashboards were made public so that everyone can access the data in not only tabular format but also as dynamically linked maps, charts and graphs. Findings reported in this paper can be independently verified on the dashboards, and additional patterns and trends may be discovered and tested on them.

Power BI and Tableau were selected among popular dashboard platforms for their functionality and performance advantages. The data preparation was done in Excel and Power Query, which allowed for the pivoting of columns into rows. The data was read as temporal and joined with Power BI's internal United States map projected in the Mercator projection. A Power BI free account was used for making and publishing the two main dashboards (2019 and 2020) after creating them on Power BI Desktop. The layout customization was first done in Figma and then the UI was transferred to

Power BI. The Tableau dashboards were made with a Tableau Desktop Student Licence. The layout of the states was provided by Brittany Fong.

The dashboards that are produced as a result of this study include the following:

(1) COVID-19 Impact on Mortality of Various Causes in the United State (with tutorial)
(2) Mortality of Various Causes in the United States – 2019 Prediction vs. Actual Comparison
(3) Predicted and Actual Deaths in the United States without COVID-19 for 2020
(4) Multiple and Underlying COVID-19 Deaths per 100,000 People in the United States for 2020

The links to access these dashboards and the video tutorial can be found on the Harvard CGA GitHub page (Kumar and Tibrewal 2021).

The first two dashboards are on Power BI. They allow the public to select certain states, causes of deaths, months and years with fully customizable queries that can be executed on the fly to see the selected data geographically, temporally, and statistically. The last two dashboards are on Tableau, which provide a space-time integrated view of monthly changes in excess deaths across the nation. These interactive query and visualization tools help make comparisons across states, seasons and causes of death quick and easy.

Results and discussion

This section evaluates the eleven causes of death and COVID-19's effect on them as seen in the dashboard. Any trend line above or below the 95% confidence interval is considered to be significant. Compared with model-predicted values from 1999 to 2019 data, most 2020 actual mortality rates increased, a few stayed stagnant, and two decreased.

Figure (2a, 2b) shows the percent of actual death values fell outside of the 95% confidence interval of model prediction in 2020 and 2019 respectively for each of the 11 causes of death, and the total death of all causes. The 'all causes' category for 2020 includes death counts listed under COVID-19 Underlying Cause of Death but does not include the counts of COVID-19 Multiple Cause of Death, because the latter is already counted under their respective main underlying causes other than COVID-19.

The COVID-19 causes did not exist prior to 2020, thus have no predicted values to compare with. In 2019, all causes have less than 6% of actual values beyond the 95% confidence interval of model prediction, either above or below. In 2020, most causes have less than 6% of actual values below the 95% confidence interval of model prediction, while most causes have much larger percent of actual values above the 95% confidence interval of model prediction, as high as 55% for the 'all causes' category, due to the excess deaths caused by COVID-19 directly and indirectly.

Across the nation, when deaths caused directly by COVID-19 are removed, deaths of all other causes combined remained at or rose above the predicted levels throughout 2020. Figure 3 shows the predicted and actual (grey and red, respectively) monthly deaths per 100,000 population across the US in 2020 after removing COVID-19 deaths. North Carolina has incomplete data for December of 2020, which caused a dip of the red line in Figure 3 that does not represent an actual decrease in death counts. Several states, such as Arizona, Missouri, Tennessee, Nevada, North Dakota and South Dakota, as well as the District of Columbia, have elevated actual deaths by other causes compared with model predictions in all months of the year. While New York, New Jersey, Massachusetts, and Connecticut have a spike in the spring coinciding with the COVID-19 deaths spike, followed by a dip in early summer, perhaps due to lockdown measures reducing exposure to other airborne diseases.

Excess death directly caused by COVID-19 has been widely reported. Our study is focused on the indirect impact of COVID-19 on excess deaths of other causes. The following section explores the patterns of the 11 causes of death in more detail.

Increases

The following four causes of death increased beyond the 95% confidence interval during the pandemic: Alzheimer's Disease, Diabetes, Unclassified Deaths, and Diseases of the Heart. This study explores the possible reasons behind the increases in Alzheimer's Disease and Diabetes deaths.

According to the dashboard, there was an increase in Alzheimer's Disease deaths in 31 states. The dashboard's graph of two states with increases are depicted in Figure 4(a,b). The increase was corroborated by the Wall Street Journal in June 2020, which reported an increase of about fifteen thousand Alzheimer's deaths nationwide during the first four months of the pandemic (Kamp and Overberg 2020). CDC's Chief of Mortality Statistics, Dr. Robert Anderson, said in June, 'It's going to take more complete data and some more time to estimate how much of these are missed COVID-19

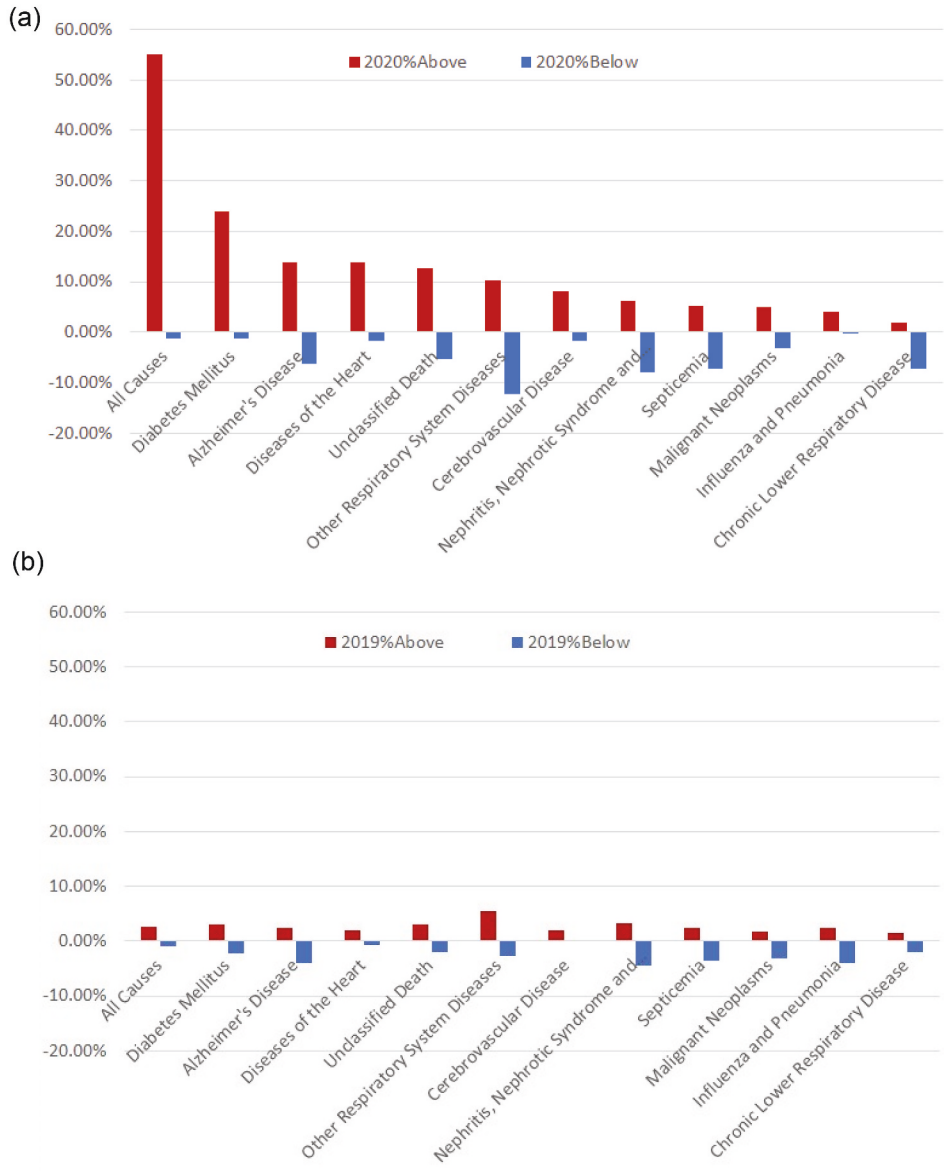

Figure 2. a. Percent of actual death values outside of the 95% confidence interval of model prediction in 2020. b. Percent of actual death values outside of the 95% confidence interval of model prediction in 2019.

deaths and how many of these are indirect.' Since that statement, the CDC has refined their data and accounted for many of these limitations.

The spike in Alzheimer's deaths may be attributed to two distinct factors. One is the shelter-in-place order. Forcing patients to be physically distant from each other and their caregivers may have caused disruption of routines in nursing homes across the country. Disruption is also one of the principal causes of dementia patients deteriorating in health (Porock et al., 2015). Loss of control often triggers disengaged and distressed behaviours in dementia patients. Many quarantined people, especially during the onset of the pandemic, reported feeling isolation, loneliness, and depression (Ettman Catherine et al. 2020). Depression is closely intertwined with dementia, with up to 68% of dementia patients also reporting traits of depression (Muliyala and Varghese 2010). The higher depression and loss of control exacerbates the effects of dementia, which might have led to more deaths.

Second, people living with Alzheimer's have a higher risk of contracting COVID-19 due to their position and exposure. Due to the patients' high genetic vulnerability to coronavirus, Alzheimer's deaths may have an increase

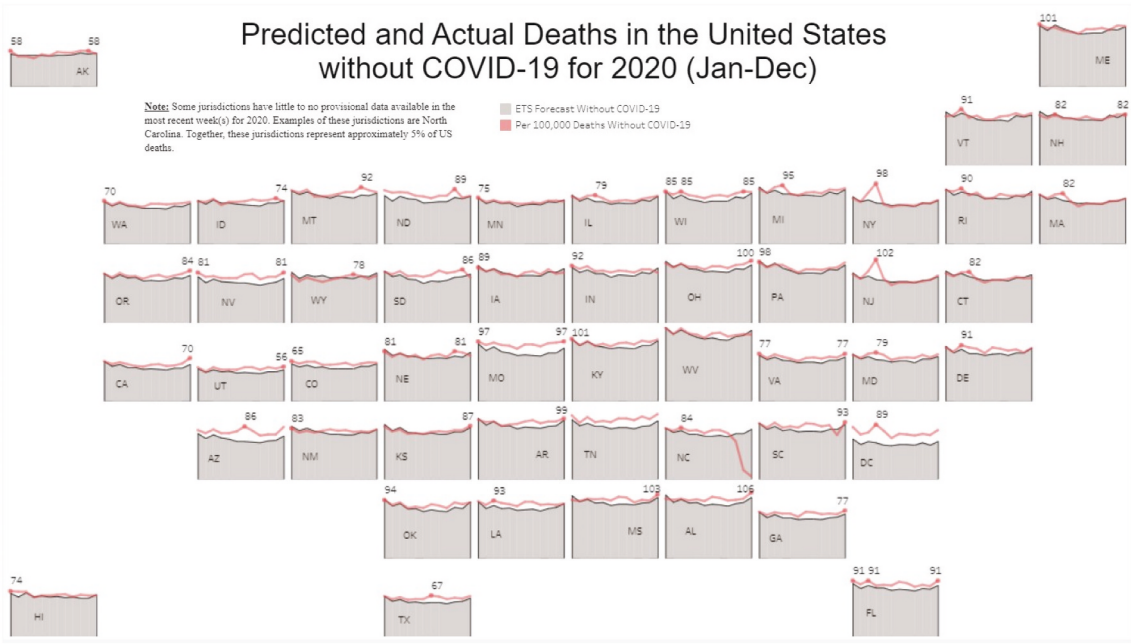

Figure 3. Predicted and actual (grey and red, respectively) monthly deaths per 100,000 population across the US in 2020 after removing COVID-19 deaths.

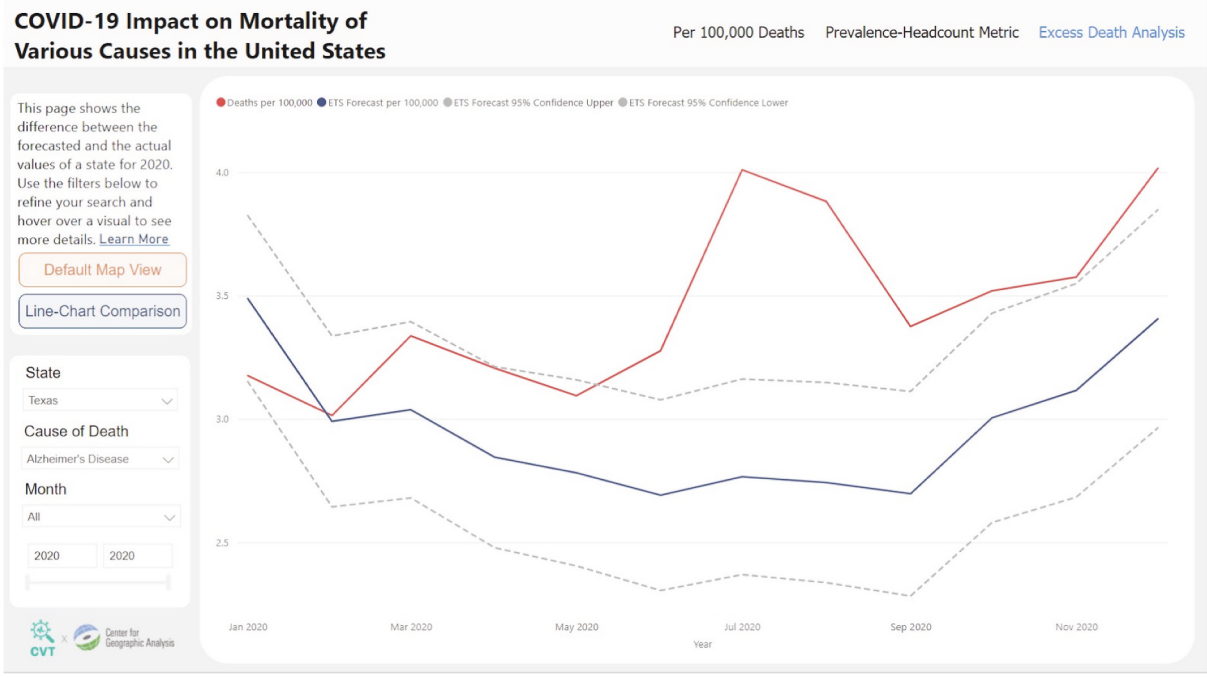

Figure 4. a. Increase in Alzheimer's Disease deaths in Texas. b. Increase in Alzheimer's Disease deaths in New York.

because they were not correctly categorized as COVID-19 deaths. Towards the beginning of the pandemic, a shortage of test kits caused many COVID-19 deaths in Alzheimer's patients to be attributed to Alzheimer's Disease. According to the New York Times in February 2021, more than 34% of all deaths due to COVID-19 occurred in nursing homes (The New York Times 2021). According to the CDC in 2016, over 40% of nursing home residents were diagnosed with forms of dementia including Alzheimer's (CDC, 2020). Therefore, those with dementia are more likely to be located in residential care communities, and those communities are highly susceptible to coronavirus deaths (Wang et al. 2021). At the early stages of the pandemic when

testing kits were limited, some of their deaths may have been counted under Alzheimer's disease rather than COVID-19.

Diabetes Mellitus, commonly known as Diabetes, also had an increase in mortality during the pandemic. According to the dashboard, there was an increase above the 95% confidence interval in Diabetes deaths during the pandemic in 38 states. The relationship between Diabetes and COVID-19 severity has been well documented since the onset of the pandemic. It is listed as one of the comorbidities of coronavirus deaths. Research by Vanderbilt University found that COVID-19 severity triples among people with Diabetes (Gregory et al. 2021). Additionally, a study scraping 61 million medical records in the United Kingdom found that 30% of COVID-19 deaths occurred in people with Diabetes (Barron et al. 2020). The correlation between Diabetes and COVID-19 mortality is due to the weakening in the immune system due to inflammation and obesity. Lower cardiorespiratory fitness among Diabetes patients makes the immune system more susceptible to the coronavirus. Just as with Alzheimer's Disease, a shortage of test kits may have caused the subsequent increase in COVID-19 deaths to be attributed incorrectly to Diabetes. The pandemic induced stress on the medical service system may have caused Diabetes patients not receiving timely and adequate care, thus increasing their death risk even if they are not infected by the coronavirus. Figure 5 shows the extent of excess death of Diabetes in Mississippi.

The remaining two causes of death that increased (Unclassified Deaths and Diseases of the Heart) are not examined in this study.

Slight increases

The two causes of death that had a slight increase in deaths were Influenza and Cerebrovascular Disease. Influenza and pneumonia, as shown by the dashboard results in Figure 6(a, b), exhibit a short season of increase in the spring, and remain at predicted levels for the rest of the year, contrary to much media portrayed decrease. The increase in the spring is likely related to insufficient testing of COVID-19. The reasoning for the expected decrease is based on the effect of COVID-19 interventions, such as lockdowns, mask-wearing, and school closures on influenza transmission, which would lessen the spread of influenza in the population . However, the CDC data does not confirm this. It is possible that the seasonal peaks of influenza (between December and February) in the US preceded the height of the pandemic, leading to no major change in influenza activity during later months (CDC 2020). Lastly, it must be emphasized that most literature surrounding seasonal influenza during the pandemic focuses on decreasing cases and not mortality (Young et al. 2020).

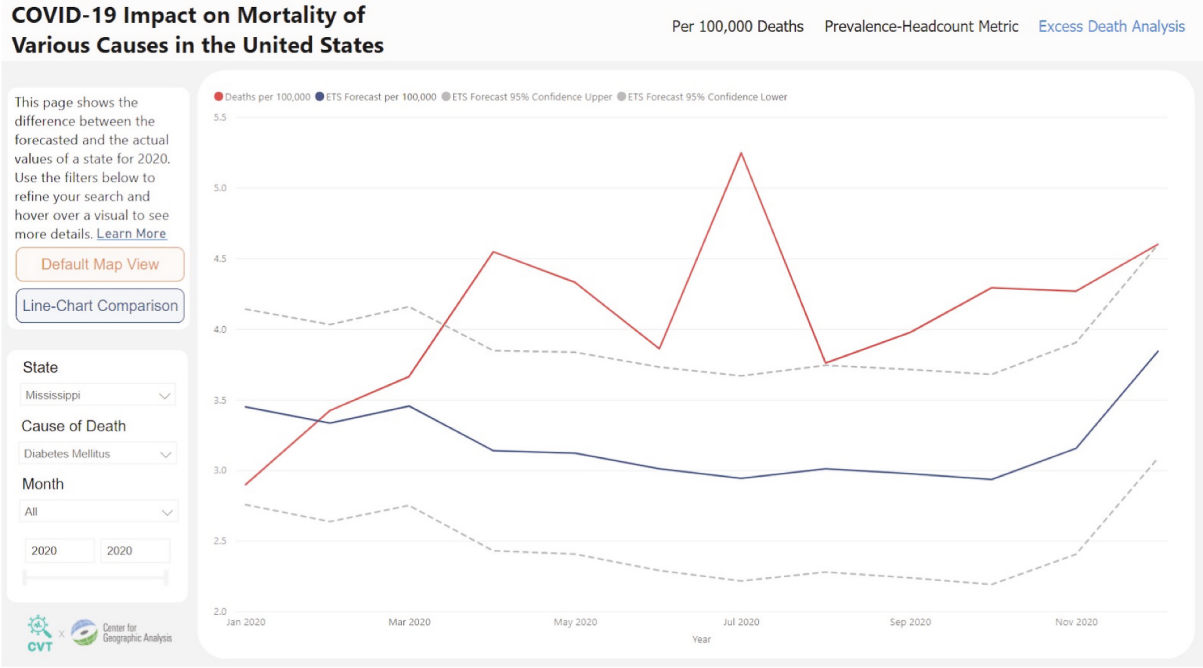

Figure 5. Increase in Diabetes deaths in Mississippi.

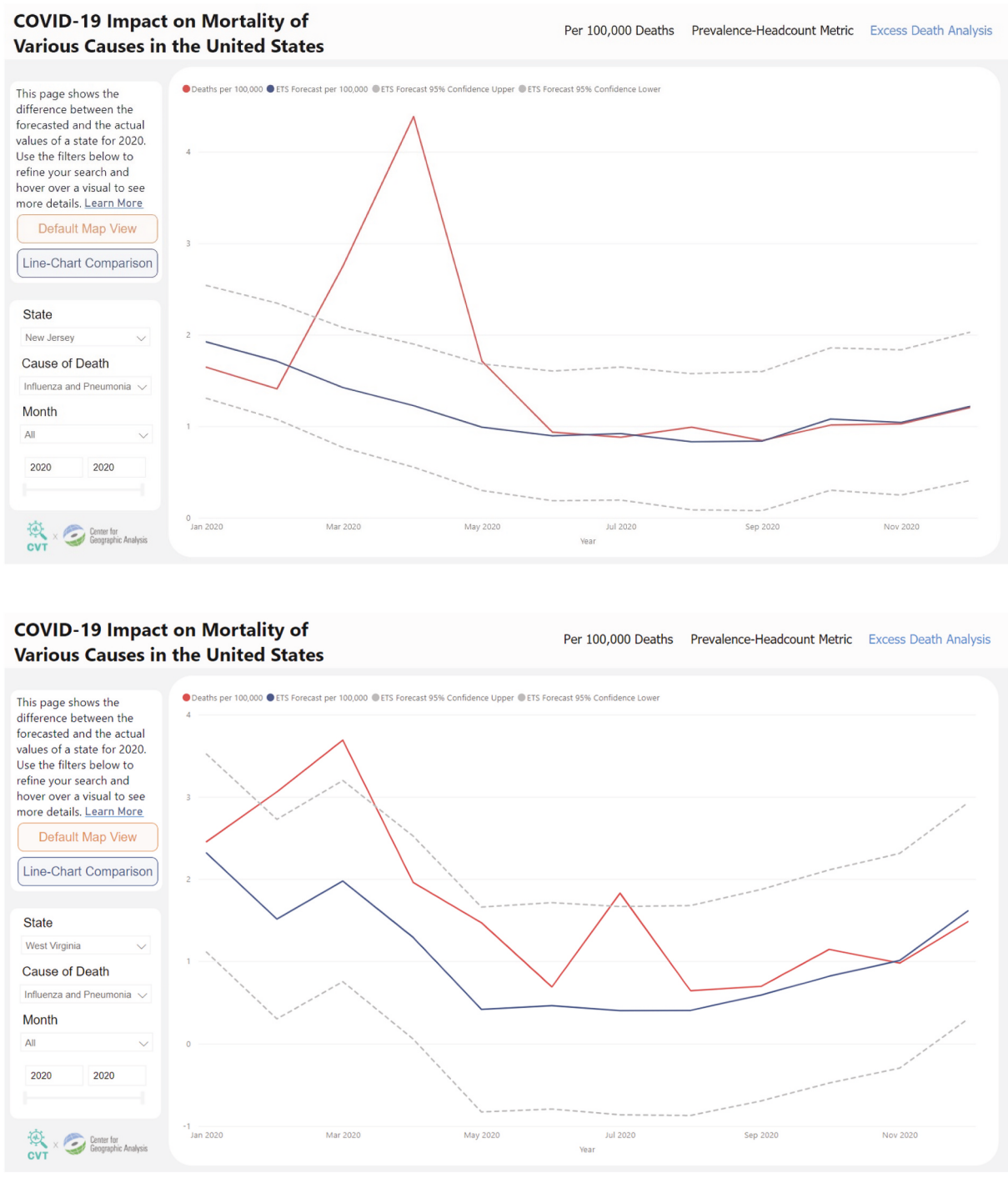

Figure 6. a. Spring spike in Influenza and Pneumonia deaths in New Jersey. b. Slight increase in Influenza and Pneumonia deaths in West Virginia.

Twenty states saw a rise above the 95% confidence interval of predictions in Cerebrovascular Disease deaths for at least one month during the pandemic, as shown in Figure 7. A literature review conducted from January through July 2020 found that there is no single pattern of cerebrovascular disease related to COVID-19. Respiratory tract infection triggers strokes and a byproduct of COVID-19, but this is not a causal relationship (Fraiman et al. 2020). Another literature review found that the relationship between Cerebrovascular Disease and COVID-19 is merely incidental (Tsivgoulis et al. 2020).

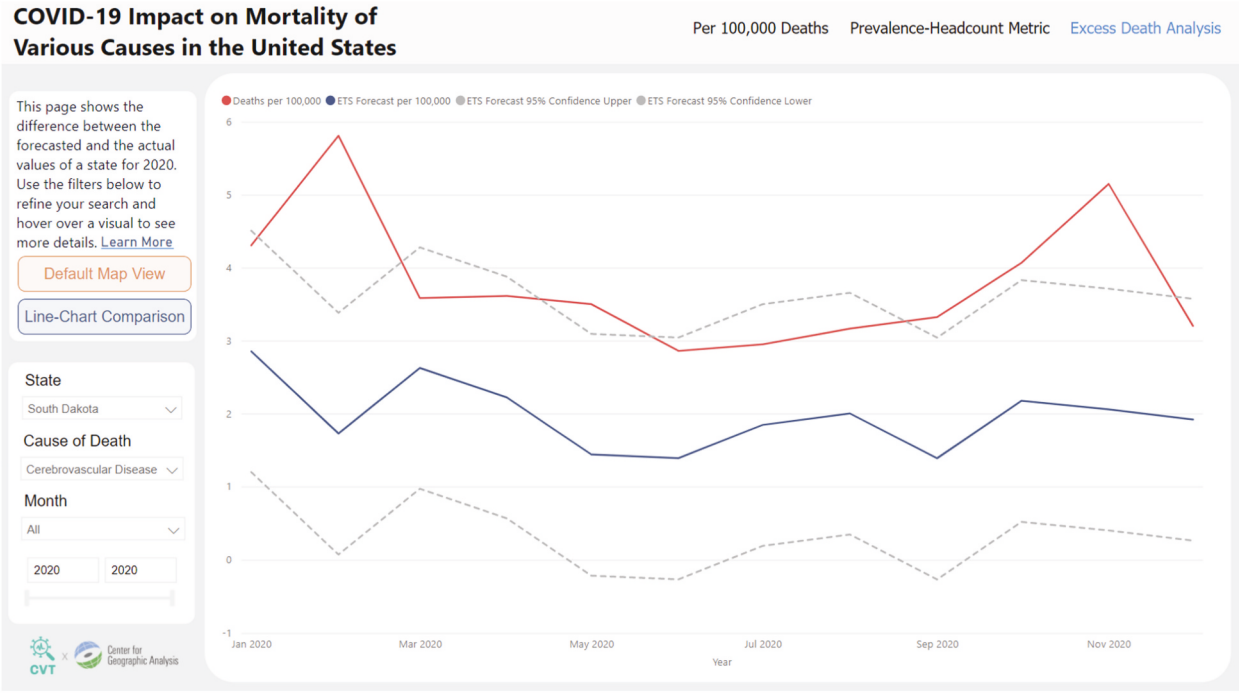

Figure 7. Slight increase in Cerebrovascular Disease deaths in South Dakota.

Decreases

Deaths of two causes decreased during the pandemic – Chronic Lower Respiratory Disease (CLRD) and Other Respiratory Diseases. According to the dashboard there was a decrease in CLRD deaths below the 95% confidence interval of model prediction in 20 states. CLRD is characterized by breathing difficulty, includes asthma, and is commonly linked to smoking (Centers for Disease Control and Prevention (US); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US) 2010). Previous research has indicated a causal relationship between a CLRD diagnosis and higher COVID-19 mortality. A meta-analysis of COVID-19 patients in China found that patients with CLRD had a 5.97-fold increased risk of requiring hospitalization due to coronavirus (Çakır Edis 2020). From this information, it would be expected that the CLRD mortality count would rise, in the same way that Alzheimer's Disease and Diabetes did due to the shortage of COVID-19 testing. In the states hit hard by the first wave, this indeed happened. The short spike was followed by a persistent decrease in CLRD deaths for the remainder of the year (Figure 8a). Other states saw a decrease in CLRD deaths through the year without the initial spike (Figure 8b).

Recent research indicates that CLRD may in fact be a positive factor during the pandemic. A study by the George Institute for Global Health found that those with asthma have a lower risk of contracting coronavirus than those without asthma and have similar clinical outcomes (Sunjaya et al. 2021). Another study found that COVID-19 patients with asthma had a lower risk of death than those without asthma (Liu et al. 2021). A study by the Boston Children's Hospital explains possible reasons for the higher resilience of asthma patients (Simoneau et al. 2020). First, better medication adherence due to coronavirus fears may have led to higher inhaler usage and lower deaths. Second, coronavirus health measures like social distancing and handwashing prevented transmission of seasonal viruses like influenza, which is a common asthma trigger. Third, environmental factors such as improved air quality due to less traffic and less contact with allergens on school playgrounds contributed to a decrease in CLRD irritants. Researchers at the University of Pennsylvania corroborated these claims, attributing the drop in CLRD deaths to social distancing, school shutdowns, and enhanced hygiene (Murez 2020). It is expected that the drop in Other Respiratory Disease deaths is due to the same reasons as the drop in CLRD deaths.

No changes

Deaths caused by Malignant Neoplasms, Nephritis, Nephrotic Syndrome and Nephrosis, and Septicaemia matched the model predictions well. This study explored the possible reasons for the trend in deaths of Malignant Neoplasms (Figure 9). Malignant Neoplasms, more

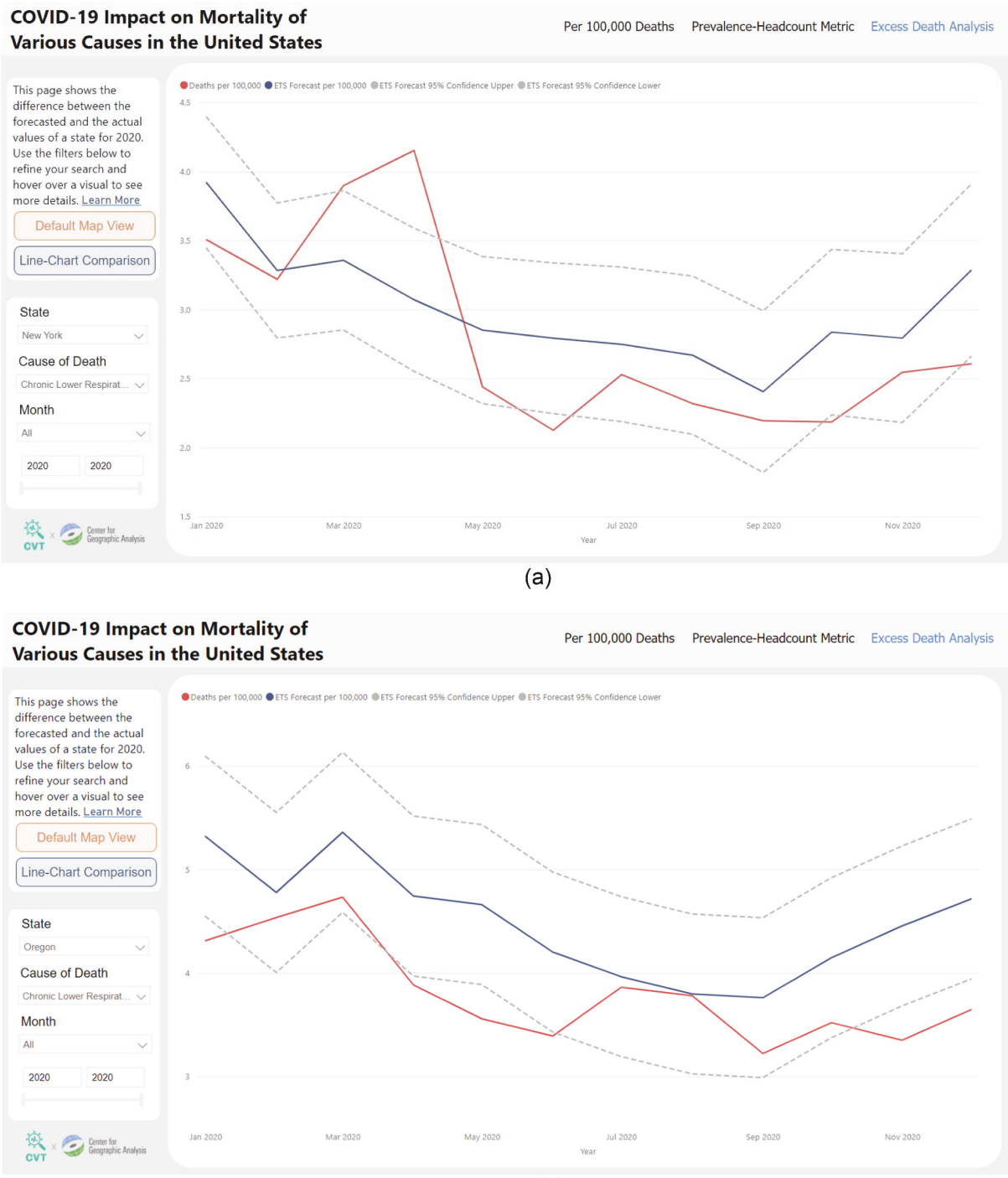

Figure 8. a. Initial spike followed by a persistent decrease in CLRD deaths in New York. b. Persistent decrease in CLRD deaths through 2020 in Oregon.

commonly referred to as cancer, are listed as a COVID-19 comorbidity. Therefore, it would be expected to increase, in line with Diabetes and Influenza. However, the dashboard indicates a stagnant count in cancer deaths, likely due to the non-emergency nature of medical care for cancer patients. Reports show that during the pandemic, chemotherapy patients are scrupulously taken care of, with excess measures taken to minimize the threat of contracting coronavirus (SSM Health 2020). Furthermore, a recent study indicated that patients receiving chemotherapy were in fact at a lower risk of contracting COVID-19 (Chen et al. 2021). Another study

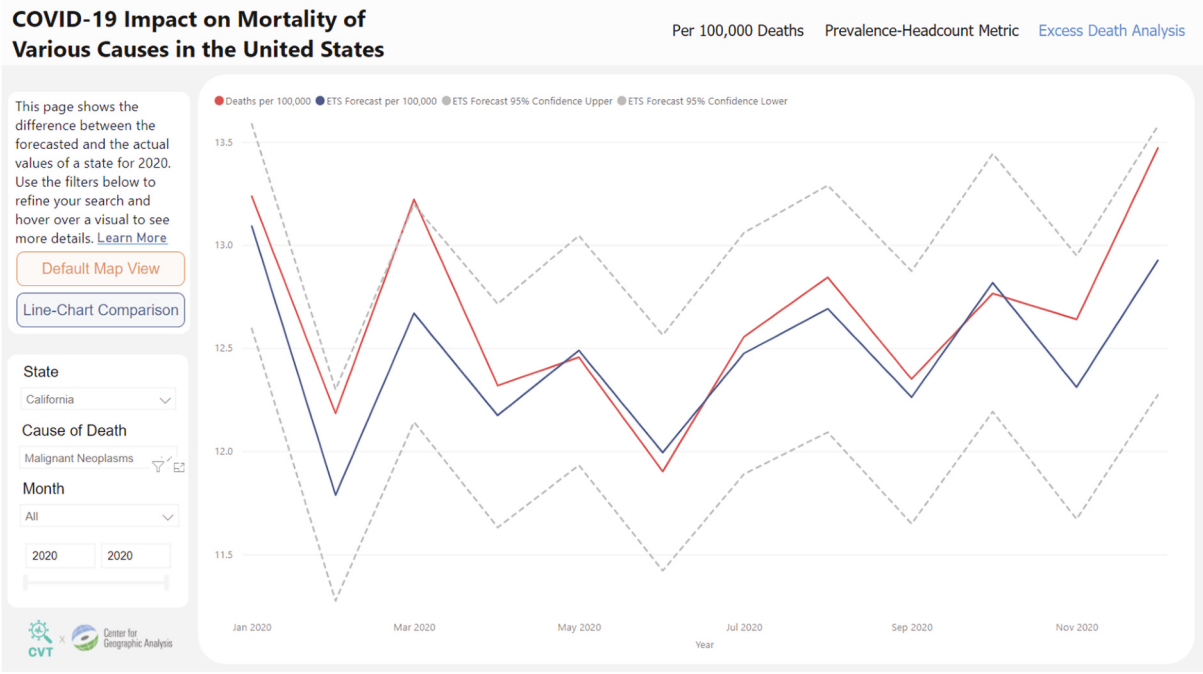

Figure 9. Comparison between actual (red) and predicted (blue) death counts of Malignant Neoplasms (Cancer) in 100,000 population in California (grey dotted lines are 95% confidence intervals of the model prediction).

found that delayed cancer treatment, was theorized to cause an increase in deaths, was found not to hinder outcomes of many patients (Fillon 2020). Malignant Neoplasms were largely isolated and did not increase its death count in accordance with COVID-19 cases.

Spatial distribution of excess death

When counting all deaths, including COVID-19 as causes, the District of Columbia, North Dakota, South Dakota, Missouri and Arizona lead the nation with more than 250 deaths per 100,000 population in excess to the model predicted values for 2020 (Figure 10).

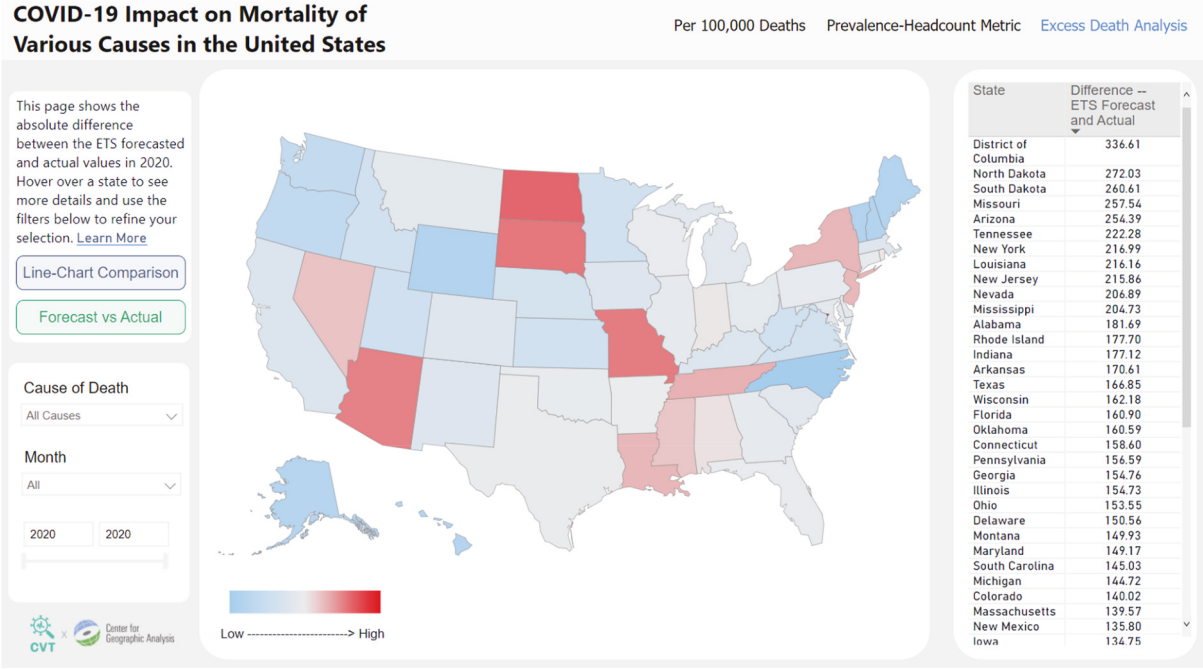

Figure 10. Spatial distribution of excess death of all causes including COVID-19 in 2020.

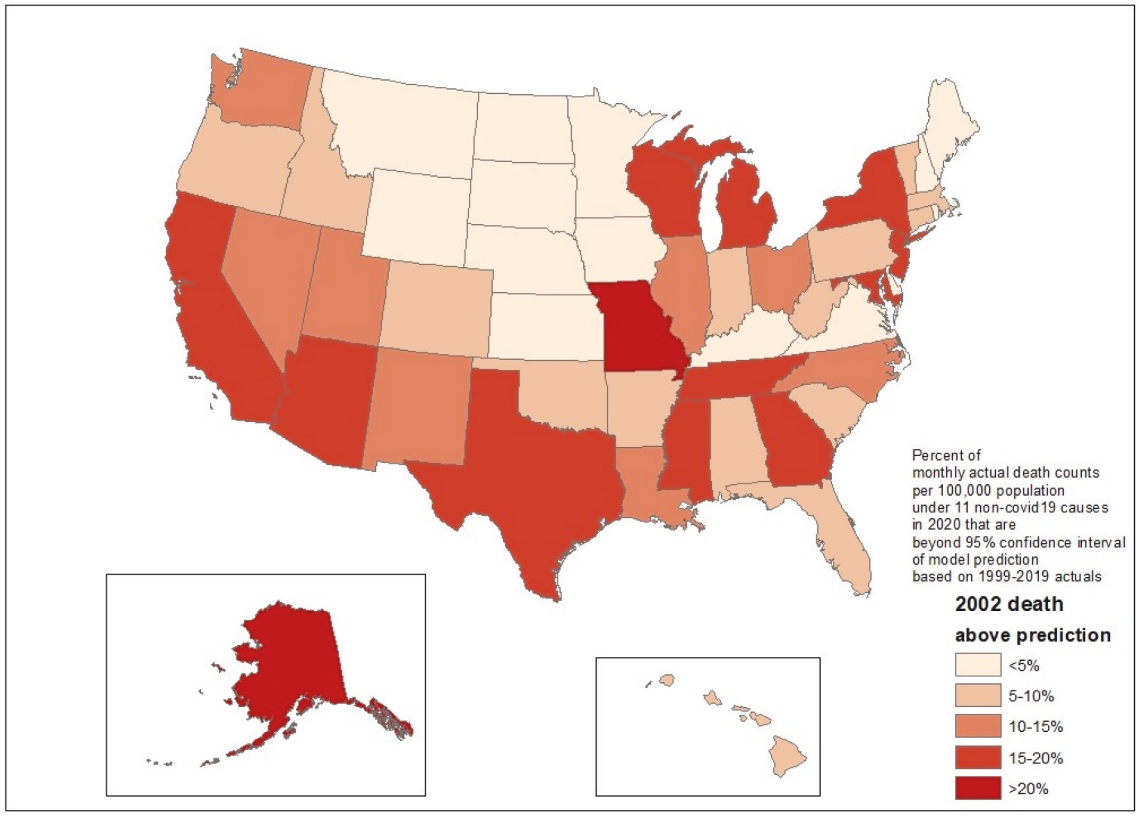

Figure 11. Percent of monthly actual death counts per 100,000 population under 11 non-COVID-19 causes in 2020 that are beyond 95% confidence interval of model prediction based on 1999–2019 actuals.

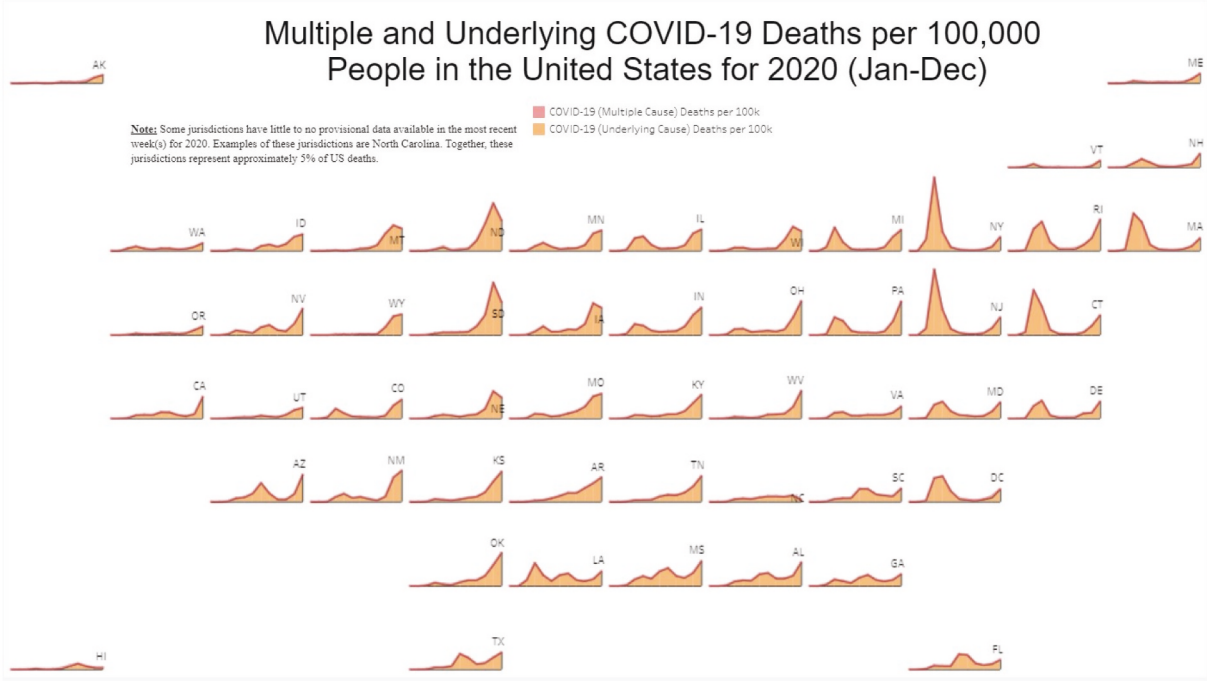

Figure 12. COVID-19 deaths across the United States through the year 2020.

When removing COVID-19 deaths, the remaining excess death of other causes shows a different pattern. Figure 11 shows the spatial distribution of excess death in 2020, excluding COVID-19 death. Monthly actual death counts per 100,000 population under 11 non-COVID-19 causes in 2020 were compared with model-predicted values based on 1999–2019 actuals, resulting in the percentage of values that are higher than the 95% confidence interval of the model prediction, which was defined as excess death. Missouri led the nation with 23% of values falling into the excess category, compared with only 1.5% for 2019. Alaska followed with 20% excess values, compared with 6.8% for 2019. Arizona, Texas, New Jersey, New York, and Tennessee have more than 17% values in the excess category. Mississippi, Georgia, Maryland, Michigan, California and Wisconsin have more than 15% in excess. Among all of them, the highest percentage for 2019 was 7.5%.

The leading states with more non-COVID-19 excess deaths are somewhat different from those states reporting the highest COVID-19 deaths. Figure 12 shows the COVID-19 death reported in 2020 by state, normalized by population and aggregated by month for comparison. Alaska is one of the states which reported the least COVID-19 deaths. Further study is needed to explain this apparent contradiction. It is possible that some states might have insufficient testing thus COVID-19 deaths were more likely miscategorized into other causes, or some states might have effective measures which curbed COVID-19 deaths but caused other stress on medical services for critically ill patients of other causes.

From Figure 12, it is evident that throughout the nation and through the year, the per 100,000 people death rates for COVID-19 being the Underlying cause are almost the same as that for COVID-19 being one of the Multiple causes. As mentioned in the Introduction section above, there is only a small percent of death (about 6%) listed COVID-19 as the Underlying cause without any other contributing causes. This study shows that there is also a small percent of death listed COVID-19 as one of the Multiple causes but not the Underlying cause. These two portions, the former included in the death counts for COVID-19 being the Underlying cause, the latter included in the death counts for COVID-19 being one of the Multiple causes, largely cancelled each other out, making the total COVID-19 Underlying and Multiple deaths very close to each other. In other words, a large majority of COVID-19 involved deaths have COVID-19 as the leading cause.

Hypothesis validation

As shown on the dashboards, findings from this study confirmed the first hypothesis. Results showed that COVID-19 death counts have not been overestimated, and the pandemic has led to more deaths among other pathologies too. COVID-19 death statistics have indeed been widely misunderstood by the general public. The main rationale for skeptics of official CDC counts is the purported misclassification of COVID-19 deaths as from other pathologies. If this were the case, we would expect COVID-19 to spike as other deaths decrease. However, we find that most causes of death did not exhibit the aforementioned behavior. On the contrary, there is an increase in death by most causes, such as Diabetes and Alzheimer's Disease, which showed synchronization in space and time with COVID-19 death. The only exceptions were the two Respiratory Disease, which exhibited a decrease, likely due to pandemic responding behaviors such as mask-wearing and lockdowns which reduced exposure to air pollutions and/or other air-borne elements of harm. As such, we can conclude that COVID-19 impact on death is not overestimated and, in fact, causes significant excess mortality in most other pathologies, in contrast to popular beliefs on social media and elsewhere.

The second hypothesis is also confirmed by the study. Our data revealed correlations between deaths by other underlying conditions and COVID-19. As stated before, we find that there is a positive correlation between spikes in deaths by Alzheimer's disease, Diabetes and COVID-19. Both Diabetes and Alzheimer's disease have been known as major comorbidities of coronavirus deaths. The excess deaths in these two causes shown by our dashboard are likely the result of misclassification of causes when COVID-19 testing was not sufficient. This is supported by data showing that both Alzheimer's Disease and Diabetes had an increase in deaths during the onset of the pandemic in New York and surrounding states, where the first wave of COVID-19 hit the hardest. Alzheimer's Disease has a genetic linkage to higher COVID-19 mortality, and the environment of nursing homes makes senior citizens with Alzheimer's more susceptible to contracting the virus. Diabetes is a COVID-19 comorbidity, with demonstrated links to respiratory complications due to higher inflammation and obesity. Therefore, the demonstrated rise in Alzheimer's Disease and Diabetes deaths may be linked to a shortage in testing, resulting in their deaths incorrectly classified as caused by those diseases alone instead of COVID-19.

The third hypothesis is that the pandemic pressure on healthcare systems accompanied by patients' fear of the virus may have led those who need medical care (for non-COVID-19 related conditions) to not be able to access it, or not receive the adequate care needed. This was especially prevalent during the beginning of the pandemic in coastal cities like New York, where healthcare systems were not prepared for COVID-19. Patients with life-threatening diseases may refrain from seeking treatment by going to the hospital out of fear of COVID-19. There has been a major decline in ED visits in Michigan for the following conditions: 23% decline for heart attack, 20% for stroke, and 10% for hyperglycaemic crisis, all of which are life threatening (Lange et al. 2020). The pandemic also caused a great shortage of medical essentials such as ventilators (CDC, 2020), Personal Protective Equipment (CDC, 2020), and blood (Stanworth et al. 2020), all leading to struggle in providing medical care. From the dashboard, it is apparent that several causes of death significantly increased throughout the pandemic in most parts of the nation. Thus, it seems that a lack of medical equipment as well as the fear of contracting COVID-19 from hospitals may have contributed to the patients not receiving adequate care, thus increasing deaths.

The fourth hypothesis is not supported by the dashboard. The Influenza and Pneumonia mortality rates were expected to decrease during the pandemic, but the dashboard indicates that this is false, as seen by the Influenza and Pneumonia fatality counts during the pandemic. On the contrary, the CDC data shows that there is a slight increase in Influenza deaths during the pandemic.

The last hypothesis is largely not supported by the data, as there are no decrease in death of most other causes, except two – the chronic lower respiratory diseases (CLRD) and the other respiratory diseases. One possible explanation for the decrease of the respiratory diseases related deaths is that the COVID-19 prevention measures also protected them from exposure to triggers of their diseases. Another possibility could be that these patients are at high risk when exposed to COVID-19. If contracted, their death would be counted under COVID-19 rather than CLRD, thus reducing the death count under CLRD alone. Further study is needed to verify such assumptions.

Limitations

There are a few limitations of this research. The first is that it is reliant on CDC definitions for causes of deaths during the pandemic. CDC defines two types of excess deaths: those having died with COVID-19 and those without. Excess death without COVID-19 is the prime focus for our research, as it could potentially include deaths indirectly caused by COVID-19 but listed under a different cause. When mentioning deaths caused by a certain disease, this research refers to the primary cause of death listed by coroners, which may or may not be misclassified.

The second limitation is that the CDC WONDER data is subject to variation in subjective judgment across jurisdictions. The National Notifiable Diseases Surveillance System uses national surveillance case definitions to try to remove as much of this subjectivity and variation as possible before the data is added to the CDC WONDER database; however, some variation may always be present.

The third limitation is the scope of our research. Due to time constraints, our research is focused on the space-time changing patterns of various deaths in relation to coronavirus. This project was unable to explore the differences of some mortality rates across states, for instance. Furthermore, our study is unfit to differentiate between various causes of the excess death. Our only finding in this regard is that some causes of death increased when COVID-19 death was high. Due to the limited spatial granularity of the data, space-time exploration and visualization are more feasible spatial analysis methods, rather than more sophisticated spatial statistical methods. CDC county level data is expected to arrive by the end of 2021, allowing for the application of spatial statistical methods such as clustering or hot spot detections, which are potential areas for further research.

The fourth limitation is the purpose of our dashboard, which was developed primarily as a data exploration and verification tool for the research team. It is open to the general public to provide succinct and factual information regarding excess COVID-19 mortality. However, this project was unable to conduct a formal dashboard development project due to resource constraints. A future area of improvement could include user requirement interviews, usability testing and functionality assessments in order to make the dashboards more professional and user-friendly.

Conclusion

The dashboards were effective in aiding the exploration of CDC data and providing readable visualizations with which to compare predicted vs actual mortality trends through 2020. The suppressed values problem was solved by estimating each cause and each state's suppressed values separately, referring to the nearest neighbouring state when the residue distribution weight is

missing. The Exponential Triple Smoothing (ETS) algorithm and a 95% confidence interval were adequate models for predicting mortality without COVID-19, providing a base to study the excess deaths caused by the pandemic.

There were four trends found in this research. First, deaths of most reported causes increased in 2020, including Diabetes, Alzheimer's, Heart Disease, and unclassified deaths. While diabetes is commonly known as a COVID-19 comorbidity, its fatality has increased independently, possibly due to reduced exercise and care, compounded by nursing home exposure without sufficient COVID-19 testing. The increase of Alzheimer's Disease deaths may be due to a combination of factors, including exposure in nursing homes without sufficient testing and depression from the shelter-in-place order.

Second, Chronic Lower Respiratory Disease and Other Respiratory Disease are the only two causes of death that showed a decrease. Contrary to media opinion, mask-wearing protocol and less exposure to air pollution may have brought the fatality rate for respiratory diseases down.

Third, influenza rates showed a slight increase, debunking a misconception that coronavirus response measures resulted in a decrease in influenza deaths.

Fourth, Malignant Neoplasms, Nephritis, Nephrotic Syndrome and Nephrosis, and Septicaemia are causes that showed no noticeable difference in death counts. A possible explanation might be that medical care for these diseases are less dependent on emergency room services compared with the other causes of death, thus less impacted by the pandemic.

The two main limitations of this study are the reliability of CDC data and the handling of suppressed values. While the CDC has made efforts during the pandemic to ensure that their handling of COVID-19 numbers is accurate, there is no guarantee that provisional data from the most recent year can withstand such scrutiny. Prior research (Faust and Del Rio 2020) indicates that data reliability, among other factors, reduces the feasibility of comparing causes of death over time. Suppressed data has a considerable impact on the prediction of each cause of death, and there has been minimal previous research on how to deal with it. This project expanded on prior methods by introducing a geographic component to the estimations.

This research has contemporary applications in statistics, public health, and policies. The project's approach to suppressed data is a novel method of data processing and may have reference value in future spatiotemporal work. The trends found from the dashboard can have impacts in US politics and in medicine, with hopefully more awareness going to common causes of death such as Diabetes and Alzheimer's Disease during the pandemic.

After completion of this paper (late March 2021), the CDC released a similar study (CDC 2021) analysing trends from 2013 to 2020 for weekly mortality data, and comparing estimated to 2020 death data in a similar fashion as this study. The conclusions of both studies agree.

Acknowledgements

This work is partially sponsored by NSF Awards #1841403 and #2027540.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation NSF [1841403, 2027540].

ORCID

Akhil Kumar http://orcid.org/0000-0001-8346-8634

Author contributions

W.G. conceived the research idea. A.K. and V.T. wrote scripts to process the data, estimated suppressed values and simulated expected death for 2020 without COVID-19. A.K. designed and constructed all dashboards. R.B. and A.C. led research on surrounding literature. Y.K. coordinated the research and writing activities. All authors contributed to writing the manuscript.

References

Barron, E., C. Bakhai, P. Kar, A. Weaver, D. Bradley, H. Ismail, and J. Valabhji. 2020. "Associations of Type 1 and Type 2 Diabetes with COVID-19-related Mortality in England: A Whole-population Study." *The Lancet Diabetes & Endocrinology* 8 (10): 813–822. doi:10.1016/S2213-8587(20)30272-2.

Baumgaertner, E., and S. Karlamangla 2020. "Chaos at Hospitals Due to Shortage of Coronavirus Testing." March 6. Accessed 17 March 2021, https://www.latimes.com/science/story/2020-03-06/chaos-at-hospitals-due-to-shortage-of-coronavirus-tests

Blue, A. 2013. Germs Spread Fast at Work, Study Finds. January 30. Accessed 17 March 2021 https://news.arizona.edu/story/germs-spread-fast-at-work-study-finds

Brown, R. B. 2020. "Public Health Lessons Learned from Biases in Coronavirus Mortality Overestimation." *Disaster Medicine and Public Health Preparedness* 14 (3): 364–371. doi:10.1017/dmp.2020.298

Çakır Edis, E. 2020. "Chronic Pulmonary Diseases and COVID-19." *Turkish Thoracic Journal* 21 (5): 345–349. doi:10.5152/TurkThoracJ.2020.20091.

Cambridge, B. D. O., MPH, and S. J. Cina MD†‡. 2010. "The Accuracy of Death Certificate Completion in a Suburban Community." *The American Journal of Forensic Medicine and Pathology* 31 (3): 232–235. September. doi:10.1097/PAF.0b013e3181e5e0e2.

CDC COVID-19 Response Team. 2020. "Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions among Patients with Coronavirus Disease 2019 - United States." February 12-March 28, 2020. MMWR. *Morbidity and mortality weekly report,* 69(13),382–386. doi:10.15585/mmwr.mm6913e2.

CDC Wonder. 2021a. "How are Small Data Values Protected to Avoid Revealing the Identities of Individual People? FAQs – Frequently Asked Questions." Accessed 5 April 2021 https://wonder.cdc.gov/wonder/help/faq.html#5

CDC Wonder. 2021b. "ICD-10 Codes, in Multiple Cause of Death 1999-2019." Accessed 5 April 2021 https://wonder.cdc.gov/wonder/help/mcd.html#ICD-10%20Codes

Center for Geographic Analysis. 2021. "Harvard Dataverse." Accessed 5 April 2021, https://dataverse.harvard.edu/dataverse/cga

Centers for Disease Control and Prevention. 2020a. "Alzheimer Disease." Accessed 17 March 2021, https://www.cdc.gov/nchs/fastats/alzheimers.htm

Centers for Disease Control and Prevention. 2020b. "*Decreased Influenza Activity during the COVID-19 Pandemic - United States, Australia, Chile, and South Africa, 2020*." Centers for Disease Control and Prevention. September 17. https://www.cdc.gov/mmwr/volumes/69/wr/mm6937a6.htm

Centers for Disease Control and Prevention. 2021. "Excess Deaths Associated with COVID-19." Centers for Disease Control and Prevention. March 10. https://www.cdc.gov/nchs/nvss/vsrr/COVID19/excess_deaths.htm

Centers for Disease Control and Prevention (US); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US). 2010. "How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General." Atlanta (GA): Centers for Disease Control and Prevention (US); 7, Pulmonary Diseases. Available from: https://www.ncbi.nlm.nih.gov/books/NBK53021/

Centers for Disease Control and Prevention, National Center for Health Statistics. 2020. "Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, Released in 2020." Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed 28 December 2020 http://wonder.cdc.gov/ucd-icd10.html 12:47:14 PM

Chen, M. F., M. T. Coronel, S. Pan, A. Kim, J. Hawley, G. Schwartz, and K. Crew. 2021. "Factors Associated with Developing COVID-19 among Cancer Patients in New York City." Clinical Cancer Research. Published March 2021. 10.1158/1557-3265.COVID-19-21-S11-02

Coleman, K. K., and W. V. Sigler. 2020. "Airborne Influenza A Virus Exposure in an Elementary School." *Scientific Reports* 10 (1): 1859. Published 2020 Feb 5. doi:10.1038/s41598-020-58588-1.

Enders, A. M., J. E. Uscinski, C. Klofstad, and J. Stoler. 2020. "The Different Forms of COVID-19 Misinformation and Their Consequences." *Harvard Kennedy School (HKS) Misinformation Review*. 10.37016/mr-2020-48

Esri. 2021, March. "About ArcGIS Pro." Accessed 17 March 2021, https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm

Ettman Catherine, K., M. Abdalla Salma, H. Cohen Gregory, V. P. M. Sampson Laura, P. M. Vivier, and S. Galea. 2020. "Galea Sandro. Prevalence of Depression Symptoms in US Adults before and during the COVID-19 Pandemic." *JAMA Network Open* 3 (9): e2019686. doi:10.1001/jamanetworkopen.2020.19686.

Fauci, A. S., H. C. Lane, and R. R. Redfield. 2020. "Covid-19 - Navigating the Uncharted." *The New England Journal of Medicine* 382 (13): 1268–1269. doi:10.1056/NEJMe2002387.

Faust, J. S., and C. Del Rio. 2020. "Assessment of Deaths from COVID-19 and from Seasonal Influenza." *JAMA Internal Medicine* 180 (8): 1045. doi:10.1001/jamainternmed.2020.2306.

Fillon, M. 2020. "Cancer Treatment Delays Caused by the COVID-19 Pandemic May Not Hinder Outcomes." *JAMA*. doi:10.3322/caac.21651.

Fraiman, P., C. Godeiro Junior, E. Moro, F. Cavallieri, and M. Zedde. 2020. "COVID-19 and Cerebrovascular Diseases: A Systematic Review and Perspectives for Stroke Management." *Frontiers in Neurology* 11: 574694. doi:10.3389/fneur.2020.574694.

Gregory, J. M., J. C. Slaughter, S. H. Duffus, T. J. Smith, L. M. LeStourgeon, S. S. Jaser, and D. J. Moore. 2021. "COVID-19 Severity Is Tripled in the Diabetes Community: A Prospective Analysis of the Pandemic's Impact in Type 1 and Type 2 Diabetes." *Diabetes Care* 44 (2): 526–532. doi:10.2337/dc20-2260.

Healthline. 2020. "Why the Flu Season Basically Disappeared This Year." Accessed 17 March 2021, https://www.healthline.com/health-news/why-the-flu-season-basically-disappeared-this-year

Human Mortality Database.2020. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org

Kamp, J., and P. Overberg 2020. "Coronavirus Pandemic Led to Surge in Alzheimer's Deaths." June 28. Accessed 17 March 2021. https://www.wsj.com/articles/coronavirus-pandemic-led-to-surge-in-alzheimers-deaths-11593345601

Kumar, A., and V. Tibrewal. 2021. "United States Excess Deaths in 2020." GitHub. https://github.com/cga-harvard/USA-Excess-Deaths

Lange, S. J., M. D. Ritchey, A. B. Goodman, T. Dias, E. Twentyman, J. Fuld, and Q. Yang. 2020. "Potential Indirect Effects of the COVID-19 Pandemic on Use of Emergency Departments for Acute Life-threatening Conditions — United States." *American Journal of Transplantation* 20 (9): 2612–2617. January–may 2020. doi:10.1111/ajt.16239.

Lee, S., Andris, C. (2020), "Excess Deaths Associated with COVID-19: County Level." Friendly Cities Lab Technical Report, Georgia Institute of Technology

Liu, S., Y. Cao, T. Du, and Y. Zhi. 2021. "Prevalence of Comorbid Asthma and Related Outcomes in COVID-19: A Systematic Review and Meta-Analysis." *The Journal of Allergy and Clinical Immunology* 9 (2): 693–701. In practice. doi:10.1016/j.jaip.2020.11.054.

Lourgos, A., and D. Heinzmann 2020. "Advocate Hospitals Pause Drive-up Coronavirus Testing, Citing National Shortage of Test Kits." March 20. Accessed 17 March 2021, https://www.chicagotribune.com/coronavirus/ct-coronavirus-testing-suspended-advocate-hospitals-20200320-2kt3g6s3dnho3ctiffx2apg3ky-story.html

Lovett, I., and D. Frosch 2020. "Tests Run Short and Results Come Slow in Coronavirus-Stricken Rural Communities." March 23. Accessed 17 March 2021, https://www.wsj.com/articles/tests-run-short-and-results-come-slow-in-coronavirus-stricken-rural-communities-11584981428

Microsoft. 2019a. "Forecasting Functions." Accessed 26 March 2021, https://support.microsoft.com/en-us/office/forecasting-functions-reference-897a2fe9-6595-4680-a0b0-93e0308d5f6e

Microsoft. 2019b. "Overview of SR-CNN Algorithm in AZURE Anomaly Detector." November 28. Accessed 26 March 2021, https://techcommunitymicrosoftcom/t5/ai-customer-engineering-team/overview-of-sr-cnn-algorithm-in-azure-anomaly-detector/ba-p/982798

Milton, D. K., M. P. Fabian, B. J. Cowling, M. L. Grantham, J. J. McDevitt, and R. A. M. Fouchier. 2013. "Influenza Virus Aerosols in Human Exhaled Breath: Particle Size, Culturability, and Effect of Surgical Masks." *PLoS Pathogens* 9 (3): e1003205. doi:10.1371/journal.ppat.1003205.

Muliyala, K. P., and M. Varghese. 2010. "The Complex Relationship between Depression and Dementia." *Annals of Indian Academy of Neurology* 13 (Suppl 2): S69–S73. doi:10.4103/0972-2327.74248.

Murez, C. 2020. "Fewer Severe Asthma Attacks In Kids During COVID." August 27. Accessed 26 March 2021, https://www.webmd.com/lung/news/20200827/one-pandemic-silver-lining-fewer-severe-asthma-attacks-in-kids#1

National Center for Health Statistics. (2021). "Weekly Counts of Deaths by State and Select Causes, 2020-2021." Accessed 25 March 2021, https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-State-and-Select-Causes/muzy-jte6

Koch, T. M. 2015. "Multiple and Underlying Causes of Death Using Montana Death Certificates, 1999-2014." Accessed 5 April 2021 https://dphhs.mt.gov/Portals/85/publichealth/documents/Epidemiology/VSU/VSU_Multiple_Cause_2017.pdf

Oransky, A. I. 2020. "*Johns Hopkins Student Newspaper Deletes, Then Retracts, Article on Faculty Member's Presentation about COVID-19 Deaths*." Retraction Watch. November 28. https://retractionwatch.com/2020/11/27/johns-hopkins-student-newspaper-deletes-then-retracts-article-on-faculty-members-presentation-about-COVID-19-deaths/

Porock, D., P. Clissett, R. H. Harwood, and J. R. Gladman. 2015. "Disruption, Control and Coping: Responses of and to the Person with Dementia in Hospital." *Ageing and Society* 35 (1): 37–63. doi:10.1017/S0144686X13000561.

Quick, H. 2019. "Estimating County-Level Mortality Rates Using Highly Censored Data From CDC WONDER." *Preventing Chronic Disease* 16: 180441. doi:10.5888/pcd16.18044.

Root, Jay. 2020. "Texans Still Clamoring for Coronavirus Tests as State Promises Availability Will Skyrocket." March 17. Accessed 17 March 2021, https://www.texastribune.org/2020/03/17/texans-still-clamoring-coronavirus-tests-state-promises-more-soon/

Schaeffer, K. 2020. "A Look at the Americans Who Believe There Is Some Truth to the Conspiracy Theory that COVID-19 Was Planned." Pew Research Center. August 18. Accessed 17 March 2021, https://www.pewresearch.org/fact-tank/2020/07/24/a-look-at-the-americans-who-believe-there-is-some-truth-to-the-conspiracy-theory-that-COVID-19-was-planned/

Simoneau, T., K. F. Greco, A. Hammond, K. Nelson, and J. M. Gaffin. 2020. "Impact of the COVID-19 Pandemic on Pediatric Emergency Department Utilization for Asthma." *Annals of the American Thoracic Society* Dec 4. Epub ahead of print. PMID: 33272107. doi:10.1513/AnnalsATS.202007-765RL.

SSM Health. 2020. "COVID-19 and Its Effect on Cancer Patients, Survivors." June 27. Accessed 30 March 2021, https://www.ssmhealth.com/blogs/ssm-health-matters/july-2020/covid-19-and-effect-on-cancer-patients-survivors

Stanworth, S. J., H. V. New, T. O. Apelseth, S. Brunskill, R. Cardigan, C. Doree, ... J. Thachil. 2020. "Effects of the COVID-19 Pandemic on Supply and Use of Blood for Transfusion." *The Lancet Haematology* 7 (10): e756–e764. doi:10.1016/s2352-3026(20)30186-1.

Subramanian, S. V., R. Sarwal, W. Joe, and R. Kim. 2020. *Geo-visualising Diet, Anthropometric and Clinical Indicators for Children in India: Enabling District Prioritisation for Interventions*. Cambridge (MA): Harvard Center for Population and Development Studies. December. New Delhi: Institute of Economic Growth. doi:10.7910/DVN/ZSH8HR.

Sunjaya, A. P., S. M. Allida, G. L. Di Tanna, and C. Jenkins. 2021. "Asthma and Risk of Infection, Hospitalisation, ICU Admission and Mortality from COVID-19: Systematic Review and Meta-analysis." *Journal of Asthma* 1–14. doi:10.1080/02770903.2021.1888116.

Susman, E. 2020. "Was COVID-19 Really the Killer?" October 15. Accessed 17 March 2021, https://www.medpagetoday.com/meetingcoverage/cap/89142

Tasnim, Samia, Md Mahbub Hossain, and Hoimonty Mazumder. 2020. "Impact of Rumors and Misinformation On COVID-19 in Social Media." Journal of Preventive Medicine and Public Health. Korean Society for Preventive Medicine. April 2. https://www.jpmph.org/journal/view.php?doi=10.3961%2Fjpmph.20.094

The New York Times. (2020). Data from The New York Times, based on reports from national and municipal health agencies. https://github.com/nytimes/COVID-19-data/tree/master/excess-deaths https://github.com/nytimes/COVID-19-data/tree/master/excess-deaths

The New York Times. 2021. More Than One-Third of U.S. Coronavirus Deaths Are Linked to Nursing Homes. February 21. Accessed 17 March 2021, https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html

The Power BI Team, Microsoft. 2014. "Describing the Forecasting Models in Power View." In Microsoft Power BI Blog. May 8. Accessed 17 March 2021, https://powerbi.microsoft.com/en-us/blog/describing-the-forecasting-models-in-power-view/

Tiwari, C., K. Beyer, and G. Rushton. 2014. "The Impact of Data Suppression on Local Mortality Rates: The Case of CDC WONDER." *American Journal of Public Health* 104 (8): 1386_1388. doi:10.2105/AJPH.2014.301900.

Tsivgoulis, G., L. Palaiodimou, R. Zand, V. A. Lioutas, C. Krogias, A. H. Katsanos, and S. Tsiodras. 2020. "COVID-19 and Cerebrovascular Diseases: A Comprehensive Overview." *Therapeutic Advances in Neurological Disorders* 13: 175628642097800. doi:10.1177/1756286420978004.

United States Mortality DataBase. (2020). University of California, Berkeley (USA). Available at usa.mortality.org

Vaziri, A. 2020. "Swab Shortage Holds up Coronavirus Testing in California." April 24. Accessed 17 March 2021, https://www.sfchronicle.com/bayarea/article/Swab-shortage-holds-up-coronavirus-testing-in-15224961.php

Vosoughi, S., D. Roy, and S. Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–1151. doi:10.1126/science.aap9559.

Wang, Q., P. B. Davis, M. E. Gurney, and R. Xu. 2021. "COVID-19 and Dementia: Analyses of Risk, Disparity, and Outcomes from Electronic Health Records in the US." *Alzheimer's & Dementia* 17 (8): 1297–1306. Published 2021 February 09. doi:10.1002/alz.12296.

Weinberger, D. M., J. Chen, T. Cohen, F. W. Crawford, F. Mostashari, D. Olson, V. E. Pitzer, et al. 2020. "Estimation of Excess Deaths Associated with the COVID-19 Pandemic in the United States, March to May 2020." *JAMA Internal Medicine* 180 (10): 1336–1344. doi:10.1001/jamainternmed.2020.3391.

Woolf, S. H., D. A. Chapman, R. T. Sabo, D. M. Weinberger, L. Hill, and D. D. H. Taylor. 2020. "Excess Deaths from COVID-19 and Other Causes, March-July 2020." *JAMA* 324 (15): 1562–1564. doi:10.1001/jama.2020.19545.

Young, G., X. Peng, A. Rebeza, S. Bermejo, C. De, L. Sharma, and C. S. Dela Cruz. 2020. "Rapid Decline of Seasonal Influenza during the Outbreak of COVID-19." *ERJ Open Research* 6 (3): 00296–2020. doi:10.1183/23120541.00296-2020.