The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: monitoring county level vulnerability

Authors: Skylar W. Marvel[1], John S. House[2], Matthew Wheeler[2], Kuncheng Song[1], Yihui Zhou[1], Fred A. Wright[1,3], Weihsueh A. Chiu[4], Ivan Rusyn[4], Alison Motsinger-Reif[2]*, David M. Reif[1]*

Affiliations:

[1] Bioinformatics Research Center, Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA.

[2] Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, 27709, USA.

[3] Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

[4] Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77845, USA.

*Correspondence to: alison.motsinger-reif@nih.gov and dmreif@ncsu.edu.

Abstract: While the COVID-19 pandemic presents a global challenge, the U.S. response places substantial responsibility for both decision-making and communication on local health authorities. To better support counties and municipalities, we integrated baseline data on relevant community vulnerabilities with dynamic data on local infection rates and interventions into a Pandemic Vulnerability Index (PVI). The PVI presents a visual synthesis of county-level vulnerability indicators that can be compared in a regional, state, or nationwide context. We describe use of the PVI, supporting epidemiological modeling and machine-learning forecasts, and deployment of an interactive, web Dashboard. The Dashboard facilitates decision-making and communication among government officials, scientists, community leaders, and the public to enable more effective and coordinated action to combat the pandemic.

One Sentence Summary: The COVID-19 Pandemic Vulnerability Index Dashboard monitors multiple data streams to communicate county-level trends and vulnerabilities and support local decision-making to combat the pandemic.

Defeating the COVID-19 pandemic requires well-informed, data-driven decisions at all levels of government, from federal and state agencies to county health departments. Numerous datasets are being collected in response to the pandemic, enabling the development of predictive models and interactive monitoring applications[1, 2]. However, this multitude of data streams—from disease incidence to personal mobility and comorbidities—is overwhelming to navigate, difficult to integrate, and challenging to communicate. Synthesizing these disparate data is crucial for decision-makers, particularly at the state and local levels, to prioritize resources efficiently, identify and address key vulnerabilities, and evaluate and implement effective interventions. To address this situation, we developed a COVID-19 Pandemic Vulnerability Index (PVI) Dashboard (https://covid19pvi.niehs.nih.gov/) for interactive monitoring that features a county-level Scorecard to visualize key vulnerability drivers, historical trend data, and quantitative predictions to support decision making at a local level (Figure 1).

We assembled U.S. county- and state-level datasets into 12 key indicators across four major domains: current infection rates (infection prevalence, rate of increase), baseline population concentration (daytime density/traffic, residential density), current interventions (social distancing, testing rates), and health and environmental vulnerabilities (susceptible populations, air pollution, age distribution, comorbidities, health disparities, and hospital beds). These 12 indicators (some of which combine multiple datasets) are integrated at the county level into an overall PVI score, employing methods previously used for geospatial prioritization and profiling[3, 4]. The individual data streams comprising these indicators measure either well-established, general vulnerability factors for public health disasters or emerging factors relevant to the COVID-19 pandemic[5]. Details of PVI score formulation and rationale are described in the Supplement, along with links to all source data. The software used to generate PVI scores and profiles from these data is freely available at https://toxpi.org. To empower additional modeling efforts, the complete time series of all daily PVI scores and data are available at https://github.com/COVID19PVI/data.

In developing the PVI, we performed rigorous statistical modeling of the underlying data to augment confidence in responsive actions, enable quantitative analysis and monitoring, and provide short-term predictions of cases and deaths. Our modeling efforts directly address the discussion in [6], by contextualizing factors such as racial differences with corrections for socioeconomic factors, health resource allocation, and co-morbidities, plus highlighting place-based risks and resource deficits that might explain spatial distributions. Specifically, three types of modeling efforts were performed and are regularly updated. First, epidemiological modeling on cumulative case- and death-related outcomes provides insights into the epidemiology of the pandemic. Second, dynamic time-dependent modeling provides similar outcome estimates as national-level models, but with county-level resolution. Finally, a Bayesian machine learning approach provides data-driven, short-term forecasts. The results of these modeling efforts are summarized below, with detailed methods and full results included in the Supplement.

With respect to factors affecting COVID-19 related mortality, we find that the proportion of Black residents and the $PM_{2.5}$ index of small-particulate air pollution are the most significant predictors among those included, reinforcing conclusions from previous reports[7]. An increase of one percentage point of Black residents is associated with a 3.3% increase in the COVID-19 death rate. The effect of a 1 g/m^3 increase in $PM_{2.5}$ is associated with an approximately 16% increase in the COVID-19 death rate, a value at the high end of a previously reported confidence interval from a report in late April 2020[7] when deaths had reached 38% of the current total.

We find that these effects persist even when including numerous additional predictors and correcting for factors such as socioeconomic status, housing density, and comorbidities. Moreover, the effects persist for a range of response values, including cumulative (i) cases, (ii) deaths, (iii) deaths as a proportion of the population, and (iv) deaths as a proportion of reported cases (Supplemental Tables 2-5). These results strongly suggest the important role of structural variation by location, which results in drastic health disparities.

A dynamic version of the generalized linear model was also built for cases and deaths as a proportion of the population to further investigate the effects of social distancing and other predictors changing daily. Final significance testing was based on bootstrapping to account for potential time-dependent correlation structures. The results (Supplemental Tables 6-7) continue to support the importance of social vulnerability indicators and quantify the impact of social distancing in the context of the model.

To accurately predict future cases and mortality, it is necessary to account for the fluid nature of the data. Accordingly, we developed a Bayesian spatiotemporal random-effects model that jointly describes the log-observed and log-death counts to build local forecasts. Log-observed cases for a given day are predicted using known covariates (e.g., population density, social distancing metrics), a spatiotemporal random-effect smoothing component, and the time-weighted average number of cases for these counts. This smoothed time-weighted average is related to a Euler approximation of a differential equation; it provides modeling flexibility while approximating potential mechanistic models of disease spread. The smoothed case estimates are used in a similar spatiotemporal model predicting future log-death counts based on a geometric mean estimate of the estimated number of observed cases for the previous seven days as well as the other data streams. The resulting county-level predictions and corresponding confidence intervals are shown (Fig. 1).

These modeling efforts support decision making in multiple ways. The epidemiological modeling will allow for testing the impact of changes in dynamic interventions, such as changes in social distancing. The forecasting efforts support short-term resource allocation decisions, such as hospital staffing or supply allocations. These forecasts also help communicate the trends that are part of the CDC's reopening criteria[8], such as whether subtle data trends translate into flattened curves. The PVI score itself constitutes an integrated indicator of vulnerability that strongly associates with mortality outcomes in the near-to-medium term. Indeed, the significant rank-correlation between daily PVI and death-related outcomes improved along a 1-, 7-, 14-, 21-, or 28-day observation period (Supplemental Figure 1). The overall PVI score highlights highly-ranked counties that should consider taking local actions or receive targeted help to mitigate undesirable outcome trends.

The interactive visualization within the PVI Dashboard is intended to communicate factors underlying vulnerability and empower community action. Figure 2 illustrates PVI for two example counties. The visualization and quantification of county-level vulnerability indicators are displayed by a radar chart, where each of the 12 indicators comprises a "slice" of the overall PVI profile. On loading, the Dashboard displays the top 250 PVI profiles (by rank) for the current day. The data, PVI scores, and predictions are updated daily, and users can scroll through historical PVI and county outcome data. Individual profiles are an interactive map layer with numerous display options/filters that include sorting by overall score, filtering by combinations of slice scores, clustering by profile similarity (i.e. vulnerability "shape"), and searching for counties by name or state (Additional functionality is detailed in the Supplement). User selection

of any county overlays the summary Scorecard and populates surrounding panels with county-specific information (Figure 1). The scrollable panels at left include plots of vulnerability drivers relative to the nation-wide distribution across all U.S. counties, with the location of the selected county delineated. The panels across the bottom of the Dashboard report cumulative county numbers of cases and deaths; timelines of cumulative cases, deaths, PVI score, and PVI rank; daily changes in cases and deaths for the most recent 14-day period (commonly used in reopening guidelines[6]; and predicted cases and deaths for a 7-day forecast horizon.

Taking all features together, the Dashboard supports the interactive evaluation and visualization of current data for localities while also providing context with respect to all U.S. counties. Full time-series of case, death, and PVI trends allow for the interrogation of the track records of counties of interest as well as the comparison of trajectories for peer counties based on differing intervention success. For example, using Orleans County (home to New Orleans, LA) as an exemplar, the multi-criteria filtering capabilities in the Dashboard were used to find a "peer county" for comparison. By bounding PVI to a similar range on vulnerability drivers ("slices") for Pop Mobility, Residential Density, and Pop Demographics, a subset of candidate counties were identified. Clayton County, GA was chosen from this clustering to illustrate the effects of dramatic differences in public action/intervention. Figure 3 shows detailed results for the two counties, which have similar baseline vulnerabilities but took divergent intervention measures at the outset of the pandemic. Specifically, pronounced differences in intervention measures (Social Distancing and Testing) is clearly associated with differing dynamics in infection rates in these counties, as visualized through the considerable differences in magnitude of the blue intervention-related) and red (infection-related) slices over time (note that all data streams are scaled so that a larger slice indicates increased vulnerability, e.g. larger blue slices represent *less* adherence to social distancing and lower testing rates). As visualized in Figure 3, the PVI rank for Orleans County improves (i.e. follows a downward rank/percentile path), effectively blunting an accelerating cases curve through early intervention. There is no such positive change for Clayton County. Thus, the PVI Dashboard enables customized empirical comparisons and evaluations across peer or comparative counties.

Numerous expert groups have coalesced around a general roadmap to address the current COVID-19 pandemic that comprises (i) reducing the spread through social distancing, (ii) gradually easing restrictions while monitoring for resurgence and healthcare overcapacity, and (iii) eventually moving to pharmaceutical interventions. However, the responsibility for navigating the COVID-19 response falls largely on state and local officials, who require data at the community level to make equitable decisions on allocating resources, caring for vulnerable sub-populations, and enhancing/relaxing social distancing measures. The goal of the COVID-19 PVI Dashboard is to empower informed actions to combat the pandemic from the local to the national level, at multiple time scales, and accomplishes this goal by combining underlying COVID-19-specific structural vulnerabilities with dynamic infection and intervention data at the county level.

Furthermore, interventions must be embraced by the general public to be effective, and interactive visualization is a proven approach to communication between diverse audiences. The PVI Dashboard provides interactive visual profiles of vulnerability atop an underlying statistical framework for comparing, clustering, and evaluating the PVI's sensitivity to component data. The county-level Scorecards illustrate both overall vulnerability and the components driving them. Finally, epidemiological and predictive modeling provides dynamic local case- and death-

related forecasts. Overall, the COVID-19 PVI Dashboard can help facilitate decision-making and communication among government officials, scientists, community leaders, and the public to enable more effective and coordinated action to combat the pandemic.

References:

1. Wynants, L., et al., *Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal.* BMJ, 2020. **369**: p. m1328.
2. ESRI. *COVID-19 GIS Hub.* 202; Available from: https://coronavirus-resources.esri.com.
3. Bhandari, S., et al., *HGBEnviroScreen: Enabling Community Action through Data Integration in the Houston-Galveston-Brazoria Region.* Int J Environ Res Public Health, 2020. **17**(4).
4. Marvel, S.W., et al., *ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models.* BMC Bioinformatics, 2018. **19**(1): p. 80.
5. Centers for Disease Control and Prevention (CDC). *Planning for an Emergency: Strategies for Identifying and Engaging At-Risk Groups*, in *A guidance document for Emergency Managers*.
6. Chowkwanyun, M. and A.L. Reed, Jr., *Racial Health Disparities and Covid-19 - Caution and Context.* N Engl J Med, 2020.
7. Wu, X., et al., *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv, 2020.
8. Centers for Disease Control and Prevention (CDC); Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/reopen-guidance.html.

Acknowledgments: We would like to thank the IT and web services staff at NIEHS for their help and support, as well as JK Cetina and DJ Reif for useful input and advice. **Funding:** This work was supported by P42 ES027704, P30 ES029067, and P30 ES025128 and by intramural funds from the National Institute of Environmental Health Sciences. **Author contributions:** *Concept and design:* SM, FW, WC, IR, AMR, DR. *Acquisition, analysis, or interpretation of data:* All authors. *Drafting of the manuscript:* All authors. *Critical revision of the manuscript for important intellectual content:* All authors. *Statistical analysis:* JH, MW, KS, YZ, AMR, DR. *Obtained funding:* WC, IR, DR, AMR. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data is available through links in the main text or the supplementary materials.

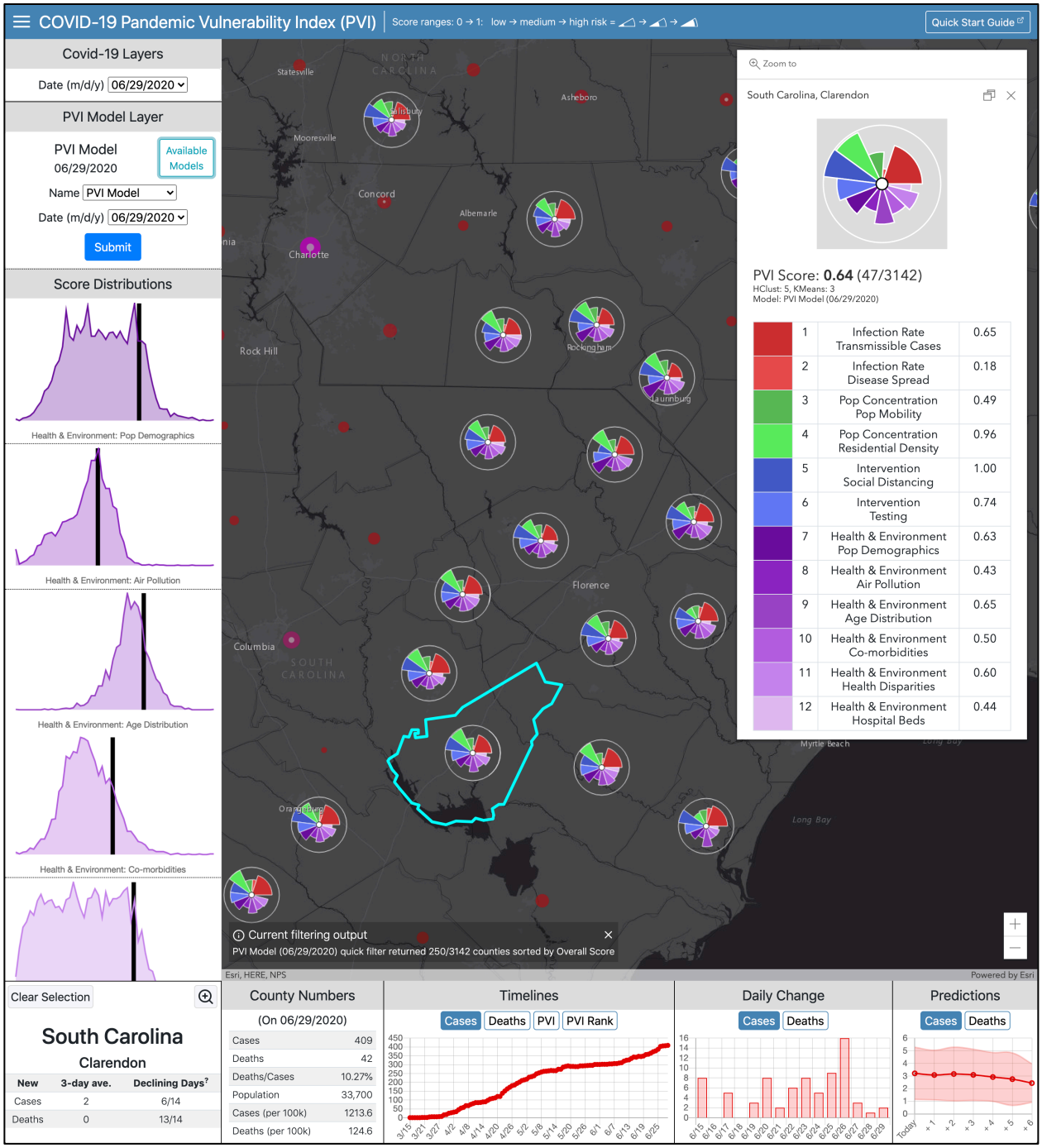

Fig. 1. Dashboard displaying map view with PVI Scorecard and associated data for a selected county. The Dashboard allows U.S.-wide navigation to area(s) of interest. The filter is set to display the top 250-ranked (i.e., most vulnerable) PVI profiles for the selected date (displayed in the upper left panels for each data layer). The Scorecard displayed shows the contribution of each indicator (slice) for Clarendon, SC, which is in a cluster of high-PVI counties across the rural Southeast U.S. The scorecard summarizes the overall PVI score and rank compared to all 3,142 U.S. counties. In the graphical profile, longer slices indicate higher vulnerability driven by a particular indicator, with corresponding indicator-wise scores (0 =

minimum; 1 = maximum) provided in the lower portion of the Scorecard. The scrollable score distributions at left compare the selected county PVI to the distributions of overall and slice-wise scores across the U.S. The panels below the map are populated with county-specific information on observed trends in cases and deaths, observed numbers for the selected date, historical timelines (for cumulative cases, cumulative deaths, PVI, and PVI rank), daily case and death counts for the most recent 14-day period, and a 7-day forecast of predicted cases and deaths. The information displayed for both observed COVID-19 data and PVI layers is scrollable back through March 2020. Documentation of additional features and usage, including advanced options (accessible via the collapsed menu at the upper left), is provided in a Quick Start Guide (linked at the upper right corner).

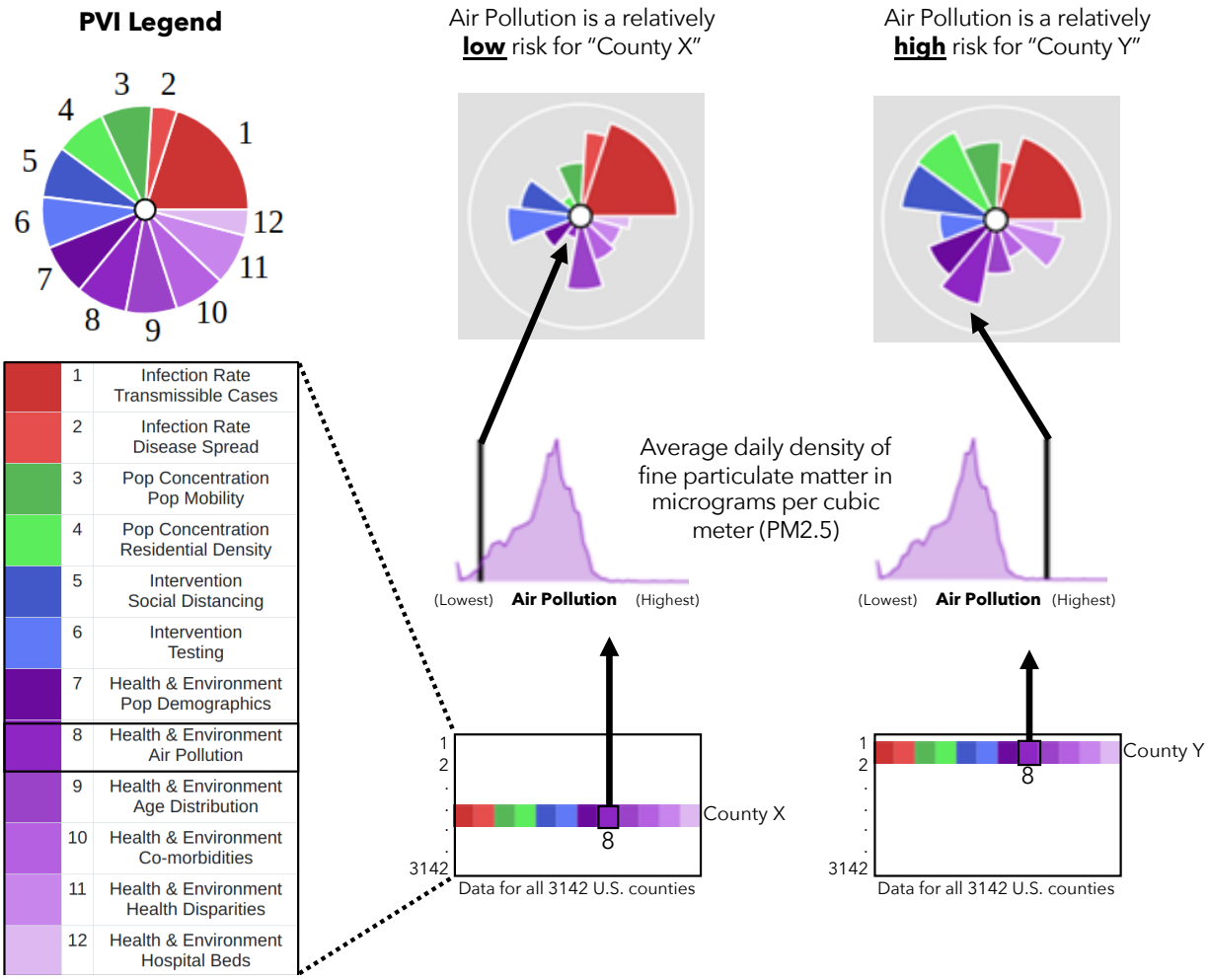

Fig. 2. Translation of data into PVI profiles. Information from all 3,142 U.S. counties is translated into PVI slices. The illustration shows how *Air Pollution* data (average density of PM2.5 per county) are compared for two example counties. The county (County Y) with the higher relative measurement has a longer *Air Pollution* slice than the county (County X) with a lower measurement. This procedure is repeated for all slices, resulting in an integrated, overall PVI profile.

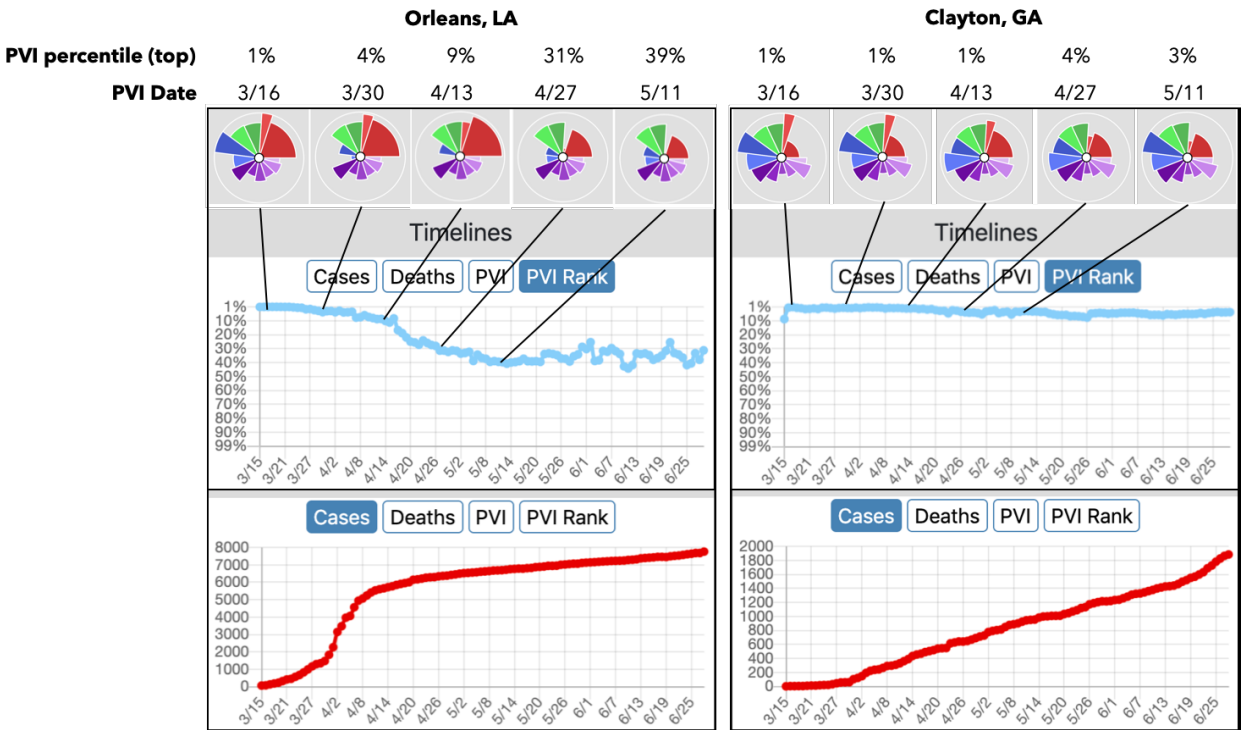

Fig. 3. PVI data from the PVI Dashboard are shown for Orleans County, LA (left) and Clayton County, GA (right). The PVI profiles for an eight-week period (from March 15th, 2020) are shown above the timelines for each county. Comparing Orleans Parish (County) to Clayton County, they have similar ranks for Population Concentration (green slices) and Health and Environment (purple slices). For Intervention Practices (blue slices), differences can be seen; Orleans County's adherence to social distancing measures and increased COVID-19 testing (reverse-scored and indicated by smaller blue slices) is visualized compared to Clayton. The changes in overall PVI rank across the timeline of the pandemic are shown. While the trajectory of new cases was blunted in Orleans County, it continued increasing in Clayton.

Supplementary Materials

Materials and Methods

Data Streams – Pandemic Vulnerability Index (PVI)

To the best of our knowledge, we have assembled the most extensive set of community-level data streams related to COVID-19 across four major domains: Infection Rate, Population Concentration, Intervention Measures, and Healthcare Vulnerability. The specific components (i.e., datasets) comprising the current Pandemic Vulnerability Index (PVI) model are provided in a dedicated Details page linked from the Dashboard. Table 1 describes each component, the rationale for its inclusion, and a link to the public data source. All source data are publicly available, and the daily assembled data are available at https://github.com/COVID19PVI/data, which is linked under Source Data in the main menu.

These data streams comprise both static and dynamic data, including static measures of population concentration and healthcare vulnerabilities. Many of the data streams are from the CDC's Social Vulnerability Index (SVI), which was developed by the Agency for Toxic Substances and Disease Registry (ATSDR's) Geospatial Research, Analysis and Services Program (GRASP). GRASP creates and maintains databases that help emergency response planners and public health officials identify and map the communities most likely to need support before, during, and after a disaster or hazard event such as the current pandemic. The SVI has been successfully used in a variety of emergency response scenarios, including mapping fire outbreaks, determining vulnerability metrics[1], and in hazard mitigation planning studies[2]. The CDC's SVI uses U.S. Census data to rank each census tract's social vulnerability based on 15 factors, including poverty, vehicle access, and housing crowding. Additional data streams from the 2020 County Health Rankings that summarize the prevalence of important co-morbidities and risk factors at the county level are also included[3]. Regarding interventions, it has been established that increased testing rates and the implementation of social distancing are effective interventions to slow the spread of COVID-19 and fatalities due to the disease[4]. Testing rates are from the COVID tracking project[5], and daily ordinal grades of social distancing adherence are calculated by Unacast[6] based on relative cell-phone mobility compared to the same period for the previous year. By anchoring movement to pre-COVID-19 activity, this measure is applicable in both rural and urban settings. Dynamic measures of disease spread and the number of transmissible cases are estimated from John Hopkins University data[7].

PVI Calculation

The PVI profiles translate numerical results into visual representations in the form of radar charts. For each county, a PVI, a dimensionless index score, is calculated as a weighted combination of all data sources and represents a formalized, rational integration of information from various domains. It is calculated using the Toxicological Prioritization Index (ToxPi) framework for data integration[8]. This integration framework has been proven to be effective for communicating risk prioritization and profiling information among scientists, regulators, stakeholders, and the general public and has been featured in publications by the U.S. National Academy of Sciences[9] and the World Health Organization's International Agency for Research on Cancer[10]. The entire set of PVI calculations can be reproduced and analyzed outside of the

Dashboard using the source data posted at https://github.com/COVID19PVI/data and in the free software application available at https://toxpi.org[11].

The PVI is visually represented as component slices of a radar chart, with each slice representing one piece (or related pieces) of information. For each profile, the radial length of a slice represents its rank relative to all other entities (i.e., counties), with a longer radius indicating higher concern or risk. The relative width (e.g., fraction of a full, 360° circle) of a slice indicates the contribution of that slice's score to the overall model. Taken as a whole, these visual profiles provide a risk assessment of the strength, relative contribution, and robustness of the multiple data sources used in the model.

PVI Weighting and Overall Vulnerability

The summarization and communication goals of the PVI and the corresponding scorecards are human-centric and designed to convey and distill high-dimensional complex data. As such, a combination of quantitative modeling and prior knowledge on risk drivers were used to inform apportionment of data into slices.

In order to gauge the association of daily PVI versus observed death outcomes, we assessed the rank-correlation between overall PVI and the key vulnerability-related outcome metrics of Cumulative Deaths (Figure 1A), Population Adjusted Cumulative Deaths (Figure 1B), and the Deaths from Cases Percentage (Figure 1C). The Spearman Rho values for PVI (from March 15 through the June 28) versus outcomes 1, 7, 14, 21, and 28 days ahead of a given day are displayed. All daily rank-correlation estimates were highly significant (all p-values < 5.1E-14). The mean Rho values typically increase with a longer time horizon (blue text on Supplementary Figures 1A,1B,1C).

Dashboard Technical Details

The Dashboard map was built using the ArcGIS JavaScript API (v4.13) and custom PHP and JavaScript files[12]. The API is used to overlay county borders, COVID-19 count data, and PVI model images on a Basemap or custom WebMap. County boundary data is obtained from a feature service while all other data is stored in an SQL database. PVI model images are base64-encoded scalable vector graphics (SVGs) rendered as inline images. The custom scripts controlling the Dashboard's functionality optimize the efficiency of HTTP requests and other computational overhead to promote real-time interactivity. The Dashboard is hosted by the National Institute of Environmental Health (NIEHS) Office of Scientific Computing, which provides high-availability HTTPS load balanced with NGINX and a secure environment for web applications. Automated data updates are pushed to the public servers daily, and the daily update process is paralleled on a private server to permit independent data integrity assessment.

Dashboard Features

Complete, continually-updated documentation is available from Dashboard links to Quick Start Guide that describes the PVI and the Dashboard tools at an introductory level (https://www.niehs.nih.gov/research/programs/coronavirus/covid19pvi/index.cfm) and a separate Details page for more in-depth information (https://www.niehs.nih.gov/research/programs/coronavirus/covid19pvi/details/). An overview of features is provided here, referring to a screenshot of the Dashboard and the expanded menu (Supplemental Figure 2).

The Dashboard displays a map of the continental U.S. by default. Users can navigate the map by dragging and dropping and by zooming in and out with the PLUS/MINUS icons or keyboard shortcuts. Clicking on a county brings up the contemporary daily scorecard along with other details. The scorecard shows the PVI graphic, the overall score, and the rank and scores for each data stream. The PVIs for the top-250 ranked counties are shown by default, and this can be changed within the quick filter panel (Panel C). Supplemental Figure 2 shows all PVI profiles for counties in the state of Alabama.

The main menu bar provides numerous options for visualizing the map. The base map can be changed from the default gray to options that include satellite imagery and topology (Change Map menu option). Additionally, users can choose a WebMap by inserting a WebMap Portal Item ID[12]. Details of the WebMap Extent options can be found through ESRI[12].

Additionally, cumulative case and death counts are plotted as a map layer default. This display option can be removed or changed through the COVID-19 Legend menu option. The size and opacity of the PVIs displayed on the map can be changed through the PVI Model Legend options. By default, the Dashboard displays the case and death counts from the Johns Hopkins Dashboard and the PVI model from the current day, and there are interactive options to display different days. Panel A (labeled in Supplemental Figure 2) allows a user to change the date for the cumulative case and death data with a scroll menu. Panel B provides similar options for the historical PVI.

Panels on the left-side and bottom of the Dashboard provide additional display options and county-level details. Panel C has several functions, all of which are accessible by scrolling down within the panel. Scrolling down the Legend panel displays the PVI Legend and the overall distribution in the form of a histogram of the overall PVI and each of the slice components across all U.S. counties. A black line indicates the location of a selected county in this distribution. For all distributions, lower values (to the left) indicate lower relative vulnerability whereas higher values (to the right) reflect higher vulnerability. Additional local information, including the county name and several metrics related to the CDC's reopening guidance, is displayed at the bottom of the Dashboard (Panel D). In addition, the number of new cases and deaths for the previous three days, as well the number of days over the last 14 days during which these numbers declined, are displayed. Both are crucial criteria in the CDC's reopening guidance[13].

Panel E displays an overall cumulative summary of the total cases, deaths, deaths/case (case fatalities), population, and per capita case and death summaries. Panel F displays the historical timeline of case counts, death counts, the PVI score, and the PVI rank. Panel G shows the daily changes in case and death counts for the previous 14 days. Panel H forecasts and confidence intervals for both cases and deaths for the following week are displayed in the bottom-right panel. Details of the predictive modeling are described below.

As shown in Supplemental Figure 2, there are several additional display options available through the main drop-down menu. In addition to the already discussed Change Map and Covid-19 Legend options, the PVI Model Filter option allows for interactive filtering for display options. With this option, the user can change the number of PVIs displayed. By default, the top-ranked counties are displayed (those with the highest vulnerability according to the overall PVI). Users can change this filtering to display the most vulnerable counties based on individual data streams, with options for multi-level filters. Additionally, options for restricting the ranges of one or more slices allow counties with similar profiles to be identified.

The PVI Model Clusters option provides additional opportunities for data-driven contextualization. Two options for clustering PVI profiles are labeled KMeans (agglomerative with k=10) and HClust (divisive, displaying the top-10 splits) after the algorithm used[14]. Both options identify counties with a similar PVI profile "shape". The distance matrix is calculated from the integrated profile that considers all slice scores and weights. For KMeans, the clustering is implemented as a Java port of the *kmeans* R function using the Hartigan-Wong algorithm[14]. For HClust, the clustering is implemented as a Java port of the *hclust* R function using the "complete" method[14]. By filtering the Dashboard to display only specific clusters, users can identify clusters with multivariate profile similarity. This can detect clusters that are geographically adjacent as well as geographically distinct to reflect the diversity of risk profiles on a national scale. Geographically disparate areas with similar drivers of concern highlight the need for integrated profiles to encourage the creation of effective targeted community policies.

Filtering and clustering options can be used iteratively to compare the histories and trajectories of similar counties, as demonstrated in the exemplar in the Report (Figure 3). These clustering and filtering tools can be used to explore the empirical, historical consequences of interventions for localities with similar baseline vulnerabilities.

The Toggle Filter Info, Toggle Dashboard, and Toggle NavBar (Navigation Bar) options allow users to customize the visualization by turning on and off the respective components on the Dashboard.

Epidemiological Modeling

While there is an ever-increasing number of models related to COVID-19, the vast array of data assembled here enables further modeling efforts. To provide context and ensure that the data streams provide conclusions and priority rankings broadly consistent with those of other groups providing epidemiological modeling, we performed cross-sectional analysis of cumulative (i) cases, (ii) deaths, (iii) deaths as a proportion of population, and (iv) deaths as a proportion of reported cases, using data current as of 6/18/2020. We emphasize that the PVI is not intended as an epidemiological modeling tool per se, e.g., it does not explicitly distinguish between aspects of vulnerability for cases vs. deaths. Our modeling here was intended to anchor the components of the PVI and to provide important context. Additionally, this modeling was not intended to provide forecasts, which are the primary focus of projection models, as discussed in the subsequent section (Functional Power Adjusted Relative Rate Modeling).

As initial analyses displayed evidence of count overdispersion, generalized linear modeling was performed using a negative binomial model with observed cumulative counts as the response, performed in R version 3.5 and the gam() procedure[14]. For analyses (i), (ii), and (iv) log(population size) values were used as predictors with estimated coefficients, and for (iii) using the 'offset' command to model deaths as a rate. Similarly, (iv) used log(cumulative cases) as an offset to model death rate among cases, an analysis which may produce biased results due to varying reporting rates by region. We note that a constant underreporting bias across counties would be absorbed into the intercept and otherwise produce valid coefficient estimates for the predictors, and this analysis may provide important clues to death risk, as by using cases in the denominator it removes a large portion of stochastic variation. Moreover, for all analyses we used the proportion tested in the state as a predictor to account for additional sources of bias.

To anchor our efforts to previous work, we included as additional fixed predictors those included in[15], which had focused primarily on the effects of a $PM_{2.5}$ air pollution index, using an analysis analogous to our model (iii). Prior to analysis, we removed predictors with pairwise

correlation greater than 0.85 with any other predictor, or where a predictor would be collinear with a series of predictors, such as overall minority proportion when individual minority groups were included. For pairs exceeding the correlation threshold, we favored the predictor with lower missingness rate (if any) or having been reported by other researchers. Dynamic predictors (i.e., those that changed substantially over the modeled period) were incorporated by using simple county averages over the March-June period covered by the PVI. With over 3100 counties (FIPS) available, most with >0 cases and deaths, the analysis could easily support the 27-28 final predictors used. To facilitate comparison with previous sources, we used predictors as given in original sources, in some instances, predictors are represented as proportions, and in other instances as percentages.

Supplementary Tables 2-7 display the regression coefficients produced with the generalized linear modeling. For cases, the most significant predictor was population size (p=3.5E-181), which might be expected. The next most significant predictors were proportion of black residents (p=2.3E-43) and mean $PM_{2.5}$ (p=2.75E-20), both associated with case counts (Supplementary Table 2). The SVI Housing measure (p=2.71E-14) and percent Hispanic (p=5.61E-11) were also associated. The proportion of population tested for SARS-Cov-2 infection was also associated (p=1.09E-10), which we attribute to statewide responses to emerging infected clusters. In this cumulative analysis, social distancing and travel-related predictors were not significant, although by mid-June 2020, many of these indexes had become more uniform than in the early days of the pandemic. For deaths as a proportion of population (Supplementary Table 4), these same predictors were highly significant, except for percent Hispanic. We note that cases and deaths per unit population represent an ultimate societal risk for which vulnerability measures are relevant, and the rank correlation coefficients for the PVI vs the fitted values for cases and death rates are 0.57 and 0.49, respectively ($p<10^{-16}$ for both).

Our analysis of deaths per cases was instructive, despite the caveats we've described in interpretation, due to potential bias due to testing variation. We note that a true case fatality rate potentially involves very different predictors than a case rate model, and deaths per unit population are also closely tied to case rates. Our modeling (Supplementary Table 5) shows that, after multiple test correction, state testing rates are no longer significant (p=0.37), which is consistent with the hypothesis that testing mainly uncovers cases, but does not predict case fatality. Proportion black residents (p=3.14E-05) and $PM_{2.5}$ (p=0.00086) remained significant, but less so than for the deaths/population model. Deaths/(reported cases) was associated with proportion of owner-occupied residences (p=9.31E-07), and inversely associated with median house value (p=0.000863). Both of these measures tend to be associated with wealth, but the relationship is complicated by the fact that higher housing prices are an impediment to home ownership.

To provide additional context, we also performed negative binomial modeling (R version 3.5 bam() with "REML" fitting)[14] of the daily cases up to 6/11/2020 (Supplementary Table 6), using the fixed county predictors as well as the dynamic predictors (not averaged). Due to the nature of the model, here we included a two-week lag cumulative number of cases as an additional predictor, as well as a smoothing spline time-dependent term to reflect a nationwide component of risk. We refer to this model as "dynamic," although it is formally a fixed effects model, treating each day outcome as an independent realization with rate determined by the predictors. To account for potential time-dependent latent correlation structures, standard errors for the coefficients were determined by bootstrapping, treating each county across all dates as an observational unit for bootstrap resampling. Bootstrapped p-values were much more strongly

significant than for the cumulative case model, which we attribute to the model's ability to account for additional sources of variation due to the use of lagged case count, the smooth time-dependent term, and the inclusion of daily dynamic predictors. Here again the most significant predictors were proportion black (p<1E-300), the two-week lag predictor (p<1E-300), statewide percent tested (p=3.28E-291), and $PM_{2.5}$ (p=9.20E-233). The analogous model was also run for deaths/(population size) (Supplementary Table 7), again with all these predictors highly significant.

In summary, the dynamic versions of the generalized linear model reinforce and amplify the conclusions from the previous cumulative models. However, the models are not designed to perform forecasting, which can be viewed as essentially a machine learning exercise, and for which careful cross-validation approaches may be used to assess accuracy.

Functional Power Adjusted Relative Rate Modeling

To predict COVID-19 case and death counts, we model the observed log-counts for new cases and deaths in each U.S. county using a Bayesian functional model. We assume deaths are a proportion of active COVID cases, and we jointly model both cases and deaths so information is shared between these observations. In this model, the expected case count is the rate offset by the observed number of deaths and is spatially modeled across the United States. Currently, the model assumes a normal distribution on the logs of the observed counts.

Let C_{tj} and D_{tj} be the observed case and death count for day t, $1 \leq t \leq T$, and county j, $1 \leq j \leq J$. For county j, we also observe a vector $x_j = \left(1, x_{1j}, \dots, x_{mj}\right)'$ of m explanatory variables that are static over time (e.g., population density) and a vector $z_{tj} = \left(z_{1tj}, \dots, z_{ktj}\right)'$ of k explanatory variables that are observed each day (e.g., social distancing metrics). Let X_j be the matrix of observed explanatory variables, both static and dynamic, for the T time points in county j. Conditional on knowledge of the rate λ_{Cij} and λ_{Dij}, we assume that C_{ij} and D_{ij} are Poisson variates where

$$\lambda_{ctj} = exp(X_j\beta + \gamma_c(s_j, t) + \log\left[\bar{C}_{tj}\right] + \epsilon_{tj}^{C}), \qquad (1)$$

and

$$\lambda_{Dtj} = exp\left(X_j\alpha + \gamma_D(s_j, t) + \log\left[\bar{\lambda}_{ctj}\right] + \epsilon_{tj}^{D}\right). \qquad (2)$$

To define these rates, $\log\left[\bar{C}_{ij}\right]$ is the log geometric-mean of the previously observed count, $\log\left[\bar{\lambda}_{ctj}\right]$ is the log of the new case rate for the last m days (i.e., $\log\left[\bar{\lambda}_{ctj}\right] = \frac{1}{m}\sum_{i=t-m}^{t} X_j\beta + \gamma_c(s_j, i) + \bar{C}_{ij}$), $\mu_c(s_j, t)$ and $\mu_D(s_j, t)$ are spatial process accounting for unobserved heterogeneity in the response at time t and county location s_j, and $\epsilon_{tj}^{C} \sim N(0, \tau_{cj}^{-1})$ and $\epsilon_{tj}^{D} \sim N(0, \tau_{Dj}^{-1})$. This defines a Poisson-lognormal model over C_{ij} and D_{ij}, which is an over-dispersed count model that allows for efficient sampling Bayesian computation through conditionally Gibbs updates.

The random fields borrow information from nearby counties under the assumption that geographically proximate counties have public health departments with similar testing strategies and testing resources and may be more similar in response, which will account for heterogeneity

not captured by the covariates. Time is included as testing strategies and testing resources are expected to change over time as well as by region.

Modeling the Spatial-Temporal Power Term

Let $\gamma_C(s_j, t) = f_C(s_j)g_C(t)$ and $\gamma_D(s_j, t) = f_D(s_j)g_D(t)$, where $f_C \sim GP(0, \sigma_C[\cdot,\cdot])$ and $f_D \sim GP(0, \sigma_D[\cdot,\cdot])$, which are Gaussian processes (GPs) [16]with a 0 mean and squared exponential covariance kernel functions $\sigma_C[s, s'] = \exp[-\tau_c(s - s')^2]$ and $\sigma_C[s, s'] = \exp[-\tau_D(s - s')^2]$, where τ_c and $\tau_D$are length-scale parameters controlling the amount of correlation between spatial locations s and s'. For f_C and f_D, it is standard to assume the covariance kernel has an unknown variance component. In our case, we fix this parameter to 1, because it is not identifiable given g_c and g_D.

For g_c and g_D, we use first-order B-splines [17], which are local linear piecewise splines. That is,

$$g_c(t) = \sum_{k=1}^{K_c} \zeta_{ck} b_{ck}(t),$$

and

$$g_D(t) = \sum_{k=1}^{K_D} \zeta_{Dk} b_{Dk}(t),$$

where $b_{ck}(t)$ and $b_{Dk}(t)$ are defined on K_c and K_D evenly spaced knots, respectively. Linear splines were chosen to minimize end-knot variability. This formulation is a tensor product formulation[17,18], which allows modeling the three-dimensional surface as the product of a two-dimensional surface and a one-dimensional surface.

Bayesian Specification and Computation

Inference and prediction were conducted using Bayes' rule. As such, all parameters in the models given in Equations (1) and (2) are given proper priors, and inference is completed using Markov Chain Monte Carlo (MCMC) methods. All coefficients on the explanatory covariates are given Normal(0,10) priors. The precision terms, that is τ_{Cj}^{-1} and τ_{Dj}^{-1} , are given Gamma(10,1) priors. For the GPs, the length-scale terms τ_c and τ_D are given discrete uniform priors over a range of equally spaced values that cover a variety of covariance functions. Finally, the ζ_{ck} and ζ_{Dk} terms, which specify the spline coefficients over the time component in the tensor product, are given Bayesian P-spline priors[19]. This allows flexible smooth modeling of the time component and defines the variance of the tensor product GP, conditional on the time component.

All priors are conditionally conjugate, and inference is conducted using Gibbs sampling. In total, 11,000 MCMC samples are taken, with the first 1,000 disregarded as burn-in. Trace plots were examined for convergence and mixing from separate chains. These indicated convergence in the chain, typically within 100 iterations, and reasonable mixing. Posterior predictive observations (i.e., future cases and deaths) are taken every 20th sample, for a total of 500 posterior predictive observations. For the Dashboard, predictions are made by taking the mean and standard deviation of these observations. Because 1 is added to the original count, 1 is subtracted from the estimate. In the rare case that the estimated data point is negative, a 0 average case count or death count is recorded. Otherwise, non-integer values are used for the forecast.

A major issue of this construction is that the computational time of the GPs dramatically increases with more observations (i.e., algorithms on covariance matrices are intrinsically $O(n^3)$). Bayesian computation using the exact GP is not feasible for 3,142 distinct geographical locations. As an alternative, we use the method developed by Moran and Wheeler[20], which uses compressed covariance matrices that are nearly exact to the original covariance matrix (i.e., constructed such that the max norm of the two matrices is approximately 1e-14), but has the added benefit of computational complexity of $O(nlog^2n)$. Utilizing these algorithms, computation time is decreased by a factor of 40, from two weeks to 6 to 8 hours. Unlike Moran and Wheeler, we do not use Ambikasaran and Darve's[21] HODLR compression technique due to its inability to scale to more than one dimension. Instead, we utilize Börm's[22] H2 matrix compression method that utilizes the H2lib matrix library.

Discussion

Our PVI model communicates an integrated concept of vulnerability that accounts for both dynamic (*Infection Rate* and *Interventions* domains) and static (community population and healthcare characteristics) drivers. In Supplemental Figure 1, the PVI is strongly associated with observed cases and deaths for a mid-term (up to one month) horizon. Thus, a highly-ranked PVI may indicate that local actions should be taken to mitigate undesirable outcomes in the future. Moreover, the data streams themselves can be used to build predictions such as ours described below. PVI visualization is a human-centric optimization that communicates how particular communities' drivers (i.e., slices) differ in terms of overall vulnerability. Presenting information in a relative sense enables the fair comparison of communities of different sizes to support prioritization decisions. Visualization promotes transparency by clarifying the judgments and trade-offs entering into an assessment while maintaining a direct link to the underlying quantitative data.

By integrating vulnerability information (both historical and forward-looking), the Dashboard supports several kinds of key decision-making for managing the ongoing pandemic. Example use cases include the priority distribution of medical resources such as hospital beds, targeted community outreach activities, and the establishment of contact-tracing mechanisms. Eventually, the PVI could be used to support the priority distribution of vaccines to highly vulnerable communities. A key utility of a public-facing, interactive dashboard is that decision-makers can point to it for support, thus promoting transparency and public buy-in for actions taken in the interest of public health. The *Timelines* panel of the Dashboard illustrates observed county-level changes over time to answer questions such as "Have we flattened the curve?" while the *Predictions* panel presents statistically robust forecasts that consider all of the data to answer the question, "What's next?".

It is clear that pandemic vulnerability and response are dynamic and differ across communities[23]. The overall goal of the COVID-19 PVI Dashboard is to provide contextualized local summaries of differential vulnerability. A growing number of individual risk factors have been highlighted in the rapidly expanding scientific studies on COVID-19. From the early identification of older individuals' vulnerability[24] to the dramatic racial disparities that have been more recently highlighted[25], there is mounting evidence that socially marginalized populations are suffering disproportionally. Unfortunately, this is not unique to the ongoing pandemic.

As recently reviewed in the New England Journal of Medicine[26] it is crucial that more data are collected and properly contextualized and analyzed to gain a comprehensive understanding of the distribution of vulnerabilities. Without sufficient explanatory context, figures on disparity can easily perpetuate harmful myths that directly undermine the goal of eliminating health disparities. For instance, myths related to biological explanations for racial health disparities, racist stereotypes about behavioral patterns, and place-based stigma can exacerbate disparities. Chowkwanyun and Reed [26] offered clear directives to help address these concerns. First, the authors called for modeling of racial differences, correcting for socioeconomic factors, health resource allocation, and co-morbidities. Second, they called for modeling that highlights place-based risks and resource deficits that may explain spatial distributions of disparities along racial lines. We directly answer these calls in our modeling efforts and demonstrate that racial disparities due to the weathering effect of systemic inequities are the largest epidemiological factor in COVID-19 vulnerability.

Unfortunately, it is clear from the resurgence of cases occurring as states reopen that containing the COVID-19 pandemic will be an ongoing battle. Monitoring will be needed for the foreseeable future, and it is generally agreed that, due to the size and diversity of the United States, reopening and closing decisions will increasingly need to be made at the local level[27]. The PVI Dashboard can be used by policymakers and decision-makers to support the inevitable prioritization of medical resources and to communicate the reasons for such decisions. While most emerging COVID-19 dashboards provide cumulative reporting measures, our Dashboard also uniquely provides several additional functions. Historic data reporting is supported, and complete timeline data is available at the county level. The PVI visualization and rank-based quantification enumerate and summarize the numerous data streams assembled. Further, the Dashboard provides links to the data assemblage to enable further research as well as data summaries related to the CDC's reopening guidance. Finally, it delivers forecasting results for short-term resource planning.

As additional data streams become available, there will be ongoing efforts to evaluate how these data should be weighted and incorporated. For example, emergent symptom tracking data could inform estimates of disease spread and unreported cases. Additionally, software development efforts are ongoing to enable the development of PVI models with other data streams, including by government agencies that have access to data that is not publicly available.

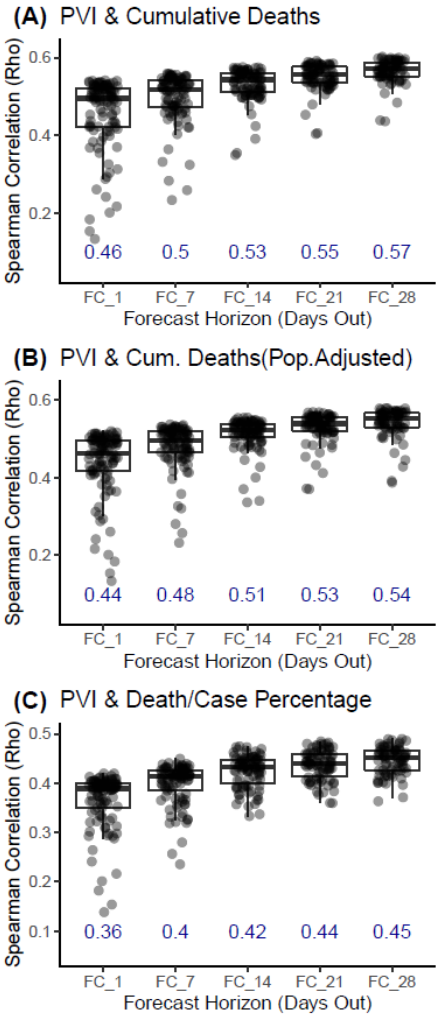

Fig. S1.

Spearman correlation (Rho) estimates of daily PVI (by county) vs. Cumulative Deaths (A), Population Adjusted Cumulative Deaths (B), and the Case Fatality Rate (C) for 1, 7, 14, 21, or 28 days into the future. Mean Rho values for each time horizon are displayed in blue.

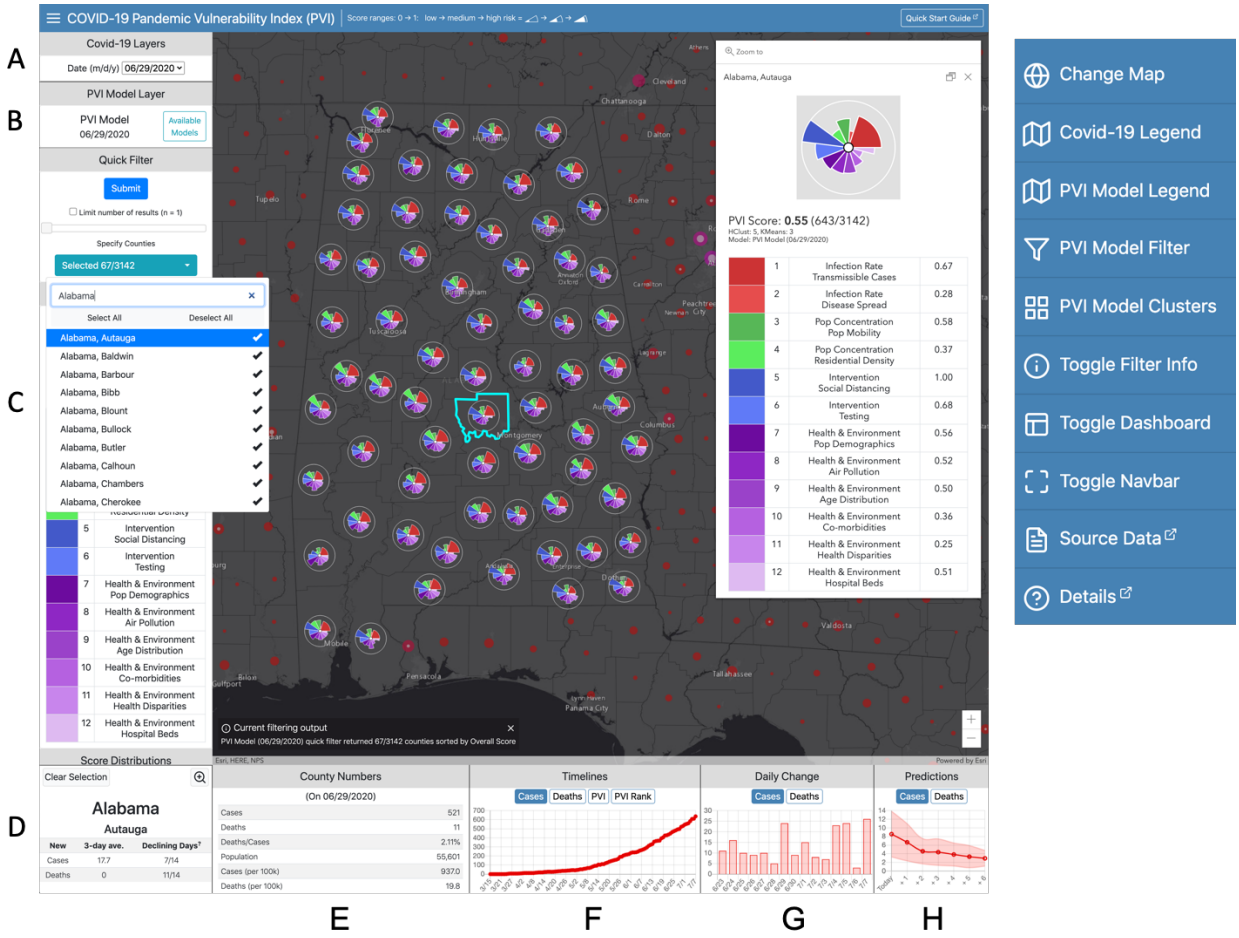

Fig. S2.

A screenshot of the Dashboard, and Menu options. Detailed descriptions of the panels and menu options are discussed in the text and available in the Quick Start Guide at https://www.niehs.nih.gov/research/programs/coronavirus/covid19pvi/index.cfm

Table S1.

Datasets comprising the current PVI model. Note that this table is available with live links on the 'Details' page at: https://www.niehs.nih.gov/research/programs/coronavirus/covid19pvi/details/

Data Domain (% weight) Data Slice (% weight) Component(s)	Update Freq.		Description/Rationale	Source(s)
Infection Rate (24%)				
Transmissible Cases (20%)				
	Daily		*Population size divided by cases from the last 14 days.* Because of the 14-day incubation period, the cases identified in that time period are the most likely to be transmissible. This metric is the number of such "contagious" individuals relative to the population, so a greater number indicates more likely continued spread of disease.	Johns Hopkins University
Disease Spread (4%)				
	Daily		*Fraction of total cases that are from the last 14 days (one incubation period).* Because COVID-19 is thought to have an incubation period of about 14 days, only a sustained decline in new infections over 2 weeks is sufficient to signal reduction in disease spread. This metric is always between 0 and 1, with values near 1 during exponential growth phase, and declining linearly to zero over 14 days if there are no new infections.	Johns Hopkins University
Population Concentration (16%)				
Population Mobility (8%)				
Daytime Population Density	Static		*Estimated daytime population.* Greater daytime population density is expected to increase the spread of infection because more people are in closer proximity to each other.	The field "DPOPDENSCY" (2019 Daytime Pop Density) from ESRI demographics analysis of American Community Survey data. 2018 CDC Social Vulnerability Index (adjunct variable)
Baseline Traffic	Static		*Average traffic volume per meter of major roadways in the county from 2018 EPA EJSCREEN.* Greater traffic volume is	2020 County Health Rankings

Data Domain (% weight) Data Slice (% weight) Component(s)	Update Freq.		Description/Rationale	Source(s)
			expected to increase the spread of infection due to more people moving and interaction.	
Residential Density (8%)				
Residential Density	Static		*Integrates data from the 2014-2018 ACS on families in multi-unit structures, mobile homes, over-crowding (more people than rooms), being without a vehicle, and persons in institutionalized group quarters.* All of these variables are associated with greater residential density, which is expected to increase the spread of infection because more people are in closer proximity to each other.	2018 CDC Social Vulnerability Index (SVI Housing Type & Transportation Theme)
Intervention Measures (16%)				
Social Distancing (8%)				
	Daily		*Unacast social distancing scoreboard grade is assigned by looking at the change in overall distance travelled and the change in nonessential visits relative to baseline (previous year), based on cell phone mobility data.* The grade is converted to a numerical score, with higher values being less social distancing (worse score) is expected to increase the spread of infection because more people are interacting with other.	Unacast
Testing (8%)				
	Daily		*Population divided by tests performed (currently only state-wide statistics are available).* This is the inverse of the tests per population, so greater numbers indicate less testing. Lower testing rates mean it is more likely that infections are undetected, so would be expected to increase the spread of infection.	The COVID tracking project
Health & Environment (44%)				
Population Demographics (8%)				
% Black	Static		*Percentage of population who self-identify as Black or African American.*	2018 "Census Population Estimates" from CHR (County Health Rankings and Roadmaps)

Data Domain (% weight) Data Slice (% weight) Component(s)	Update Freq.		Description/Rationale	Source(s)
% Native	Static		*Percentage of population who self-identify as American Indian or Alaska Native.*	2018 "Census Population Estimates" from CHR (County Health Rankings and Roadmaps)
Air Pollution (8%)				
	Static		*Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5) from 2014 Environmental Public Health Tracking Network.* Air pollution has been associated with more severe outcomes from COVID-19 infection.	M 2.5 - 2014 "Environmental Public Health Tracking Network"
Age Distribution (8%)				
% age 65 and over	Static		*Aged 65 or Older from 2014-2018 ACS.* Older ages have been associated with more severe outcomes from COVID-19 infection.	2018 CDC Social Vulnerability Index
Co-morbidities (8%)				
Premature death	Static		*Years of potential life lost before age 75 per 100,000 population (age-adjusted) based on 2016-2018 National Center for Health Statistics - Mortality Files.* This is a broad measure of health, and a proxy for cardiovascular and pulmonary diseases that have been associated with more severe outcomes from COVID-19 infection.	2020 County Health Rankings
Smoking	Static		*Percentage of adults who are current smokers from 2017 Behavioral Risk Factor Surveillance System.* Smoking has been associated with more severe outcomes from COVID-19 infection, and also causes cardiovascular and pulmonary disease.	2020 County Health Rankings
Diabetes	Static		*Percentage of adults aged 20 and above with diagnosed diabetes from 2016 United States Diabetes Surveillance System.* Diabetes has been associated with more severe outcomes from COVID-19 infection.	2020 County Health Rankings
Obesity	Static		*Percentage of the adult population (age 20 and older) that reports a body mass index (BMI) greater than or equal to 30 kg/m2.* Obesity has been associated with more severe outcomes from COVID-19 infection.	2020 County Health Rankings

Data Domain (% weight) Data Slice (% weight) Component(s)	Update Freq.		Description/Rationale	Source(s)
Health disparities (8%)				
Uninsured	Static		*Percentage uninsured in the total civilian noninstitutionalized population estimate, 2014- 2018 ACS.* Individuals without insurance are more likely to be undercounted in infection statistics, and may have more severe outcomes due to lack of treatment.	2018 CDC Social Vulnerability Index (adjunct variable)
SVI Socioeconomic Status	Static		*Integrates data from 2014-2018 ACS on percent below poverty, percent unemployed (historical), income, and percent without a high school diploma.* Lower SES are more likely to be undercounted in infection statistics, and may have more severe outcomes due to lack of treatment.	2018 CDC Social Vulnerability Index (SVI Socioeconomic Status score)
Hospital Beds (4%)				
	Static		*Summation of hospital beds for hospitals with "OPEN" status and "GENERAL MEDIAL AND SURGICAL" description.*	Homeland Infrastructure Foundation-Level Data (HIFLD)

Supp. Table 2. Coefficient table for negative binomial model with cumulative cases as response. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**StdError**	**Z**	**P-value**
(Intercept)	-7.6613	0.500325123	-15.31	6.29E-53*
AvgSocialDist	-0.0263	0.041219080	-0.64	0.523584
AvgStatePctTested	15.1075	2.340585167	6.45	1.09E-10*
CHR_Diabetes	0.4373	0.585075480	0.75	0.454816
CHR_PrematureDeath	0.000071	0.000012420	-5.72	1.05E-08*
CHR_Smoking	0.9543	0.993469677	0.96	0.336785
CHR_Traffic	0.00031	0.000103312	3.03	0.002423
education	-0.2629	0.259135057	-1.01	0.310393
hispanic	1.3958	0.212983055	6.55	5.61E-11*
INSURANCE_PctTot	0.0150	0.005227412	2.88	0.004008
log(SVI_TotPop)	0.9044	0.031508901	28.70	3.5E-181*
mean_pm25	0.1194	0.012941661	9.23	2.75E-20*
medhouseholdincome	0.0000063	0.000002759	2.30	0.021491
medianhousevalue	0.00000034	0.000000340	1.00	0.319432
pct_asian	-3.6985	1.526069013	-2.42	0.015371
pct_blk	2.3851	0.172743657	13.81	2.3E-43*
pct_native	2.1006	0.402934234	5.21	1.86E-07*
pct_owner_occ	0.8420	0.309372639	2.72	0.006493
popdensity	0.000030	0.000032788	0.90	0.370002
poverty	0.1762	0.419235332	0.42	0.6743
q_popdensity	0.0387	0.026782951	1.45	0.148235
SVI_1_Socioeconomic	0.00088	0.000993633	0.89	0.374366
SVI_2_Household	-0.2138	0.098976266	-2.16	0.030784
SVI_4_Housing	0.6783	0.089107272	7.61	2.71E-14*
SVI_AreaSqMiles	-0.000014	0.000016977	-0.81	0.42064
SVI_PctGE65	-0.0329	0.007706164	-4.27	1.95E-05*
SVI_PctLE17	0.0403	0.009333816	4.32	1.56E-05*
TRAVEL_DaytimePopDensity	-0.0000052	0.000011407	-0.45	0.649221

Supp. Table 3. Coefficient table for negative binomial model with cumulative deaths as response. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**StdError**	**Z**	**P-value**
(Intercept)	-14.5857	0.755580922	-19.30	4.97E-83*
AvgSocialDist	-0.2003	0.060312992	-3.32	0.000896*
AvgStatePctTested	18.5530	3.306836572	5.61	2.02E-08*
CHR_Diabetes	0.4996	0.892805437	0.56	0.575732
CHR_PrematureDeath	-0.0000080	0.000019905	-0.40	0.68667
CHR_Smoking	-2.4213	1.468211632	-1.65	0.099114
CHR_Traffic	0.00016	0.000136015	1.20	0.231292
education	-0.3461	0.400315979	-0.86	0.387345
hispanic	0.7066	0.329763988	2.14	0.032138
INSURANCE_PctTot	0.0025	0.008043211	0.31	0.755305
log(SVI_TotPop)	1.0628	0.045861362	23.17	8.2E-119*
mean_pm25	0.1573	0.018770331	8.38	5.18E-17*
medhouseholdincome	0.000020	0.000004086	4.81	1.51E-06*
medianhousevalue	-0.00000067	0.000000509	-1.31	0.191345
pct_asian	-4.6274	2.050886408	-2.26	0.024054
pct_blk	3.2938	0.243545905	13.52	1.12E-41*
pct_native	3.0556	0.584757770	5.23	1.74E-07*
pct_owner_occ	2.1070	0.469486930	4.49	7.19E-06*
popdensity	0.00010	0.000043146	2.38	0.01715
poverty	0.2171	0.656110351	0.33	0.740769
q_popdensity	0.1266	0.039126183	3.23	0.001218*
SVI_1_Socioeconomic	0.0018	0.001627372	1.11	0.26889
SVI_2_Household	0.2323	0.150213614	1.55	0.122048
SVI_4_Housing	0.5432	0.130983866	4.15	3.36E-05*
SVI_AreaSqMiles	-0.0000032	0.000025078	-0.13	0.896928
SVI_PctGE65	0.0062	0.011665122	0.53	0.597271
SVI_PctLE17	0.0249	0.014530061	1.72	0.086234
TRAVEL_DaytimePopDensity	-0.000028	0.000014917	-1.91	0.056399

Supp. **Table 4**. Coefficient table for negative binomial model with cumulative death rate (proportion per unit population) as response, and thus population is not included as a predictor. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**StdError**	**Z**	**P-value**
(Intercept)	-14.2003	0.699182940	-20.31	1.05E-91*
AvgSocialDist	-0.1615	0.053862659	-3.00	0.002711
AvgStatePctTested	19.0089	3.297520189	5.76	8.19E-09*
CHR_Diabetes	0.5149	0.891270015	0.58	0.563476
CHR_PrematureDeath	-0.0000092	0.000019823	-0.47	0.640832
CHR_Smoking	-2.4087	1.466219814	-1.64	0.10042
CHR_Traffic	0.00019	0.000135175	1.37	0.169637
education	-0.3756	0.398337554	-0.94	0.345783
hispanic	0.7497	0.326453187	2.30	0.021653
INSURANCE_PctTot	0.0012	0.008008757	0.16	0.876113
log(SVI_TotPop)	0.0000	0.000000000	0.00	0.00
mean_pm25	0.1570	0.018747181	8.37	5.55E-17*
medhouseholdincome	0.000020	0.000004076	4.85	1.26E-06*
medianhousevalue	-0.00000053	0.000000505	-1.05	0.292219
pct_asian	-4.5682	2.041647563	-2.24	0.025253
pct_blk	3.3255	0.242020506	13.74	5.79E-43*
pct_native	3.1182	0.582844196	5.35	8.8E-08*
pct_owner_occ	2.1288	0.467294605	4.56	5.23E-06*
popdensity	0.00012	0.000041673	2.84	0.00455
poverty	0.1633	0.655037186	0.25	0.803127
q_popdensity	0.1496	0.034073199	4.39	1.13E-05*
SVI_1_Socioeconomic	0.0019	0.001626045	1.15	0.251568
SVI_2_Household	0.2229	0.149714685	1.49	0.13659
SVI_4_Housing	0.5681	0.128992166	4.40	1.06E-05*
SVI_AreaSqMiles	0.0000067	0.000023598	0.28	0.77726
SVI_PctGE65	0.0061	0.011643401	0.53	0.599168
SVI_PctLE17	0.0253	0.014488812	1.74	0.081264
TRAVEL_DaytimePopDensity	-0.000033	0.000014567	-2.27	0.023359

Supp. Table 5. Coefficient table for negative binomial model with cumulative death rate (proportion among cases) as response. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**StdError**	**Z**	**P-value**
(Intercept)	-6.4662	0.561569619	-11.51	1.12E-30*
AvgSocialDist	-0.1383	0.046461338	-2.98	0.002912
AvgStatePctTested	4.9826	2.393042492	2.08	0.03733
CHR_Diabetes	-0.7654	0.716215789	-1.07	0.285192
CHR_PrematureDeath	0.000033	0.000016014	2.07	0.03826
CHR_Smoking	-0.3362	1.145636705	-0.29	0.769188
CHR_Traffic	-0.000034	0.000087962	-0.38	0.701985
education	-0.3051	0.318801406	-0.96	0.338558
hispanic	-0.4581	0.267575366	-1.71	0.086871
INSURANCE_PctTot	-0.0032	0.006291634	-0.51	0.610505
log(SVI_TotPop)	0.0994	0.033500694	2.97	0.003009
mean_pm25	0.0470	0.014109635	3.33	0.00086*
medhouseholdincome	0.000011	0.000003225	3.26	0.001131*
medianhousevalue	-0.0000013	0.000000390	-3.33	0.000863*
pct_asian	0.7393	1.368349412	0.54	0.589015
pct_blk	0.7451	0.178966072	4.16	3.14E-05*
pct_native	1.0125	0.459637895	2.20	0.027609
pct_owner_occ	1.7578	0.358311887	4.91	9.31E-07*
popdensity	0.000048	0.000028210	1.70	0.090013
poverty	-0.0898	0.522380987	-0.17	0.863503
q_popdensity	0.0871	0.030063029	2.90	0.003782
SVI_1_Socioeconomic	0.00082	0.001284889	0.64	0.524168
SVI_2_Household	0.2354	0.117542676	2.00	0.045235
SVI_4_Housing	0.0763	0.097153656	0.79	0.43222
SVI_AreaSqMiles	-0.0000026	0.000018696	-0.14	0.890862
SVI_PctGE65	0.0234	0.008690582	2.69	0.00711
SVI_PctLE17	-0.0244	0.011454379	-2.13	0.033
TRAVEL_DaytimePopDensity	-0.000013	0.000009687	-1.33	0.184995

Supp. Table 6. Coefficient table for negative binomial model with daily cases as response, up to 6/11/2020, using both fixed and dynamic predictors, and standard errors evaluated by bootstrapping. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**Bootstrap SE**	**Z**	**P-value**
(Intercept)	-1.549E+01	0.133877345	-115.6799302	<1E-300
AvgSocialDist	-4.661E-02	0.008746657	-5.329085351	9.87E-08*
AvgStatePctTested	1.953E+01	0.535384366	36.47029901	3.28E-291*
CHR_Diabetes	-8.227E-02	0.131723469	-0.624569838	5.32E-01
CHR_PrematureDeath	-6.666E-05	2.97096E-06	-22.43821283	1.67E-111*
CHR_Smoking	1.449E+00	0.210750288	6.876838433	6.12E-12*
CHR_Traffic	2.215E-04	1.85063E-05	11.97142093	5.02E-33*
education	-5.825E-02	0.057982156	-1.004595207	3.15E-01
hispanic	7.393E-01	0.046809115	15.79383069	3.43E-56*
INSURANCE_PctTot	1.283E-02	0.001159914	11.0647858	1.86E-28*
log(SVI_TotPop)	8.436E-01	0.006517984	129.4298013	<1E-300*
mean_pm25	8.704E-02	0.002672048	32.57521697	9.20E-233*
medhouseholdincome	1.002E-05	5.83275E-07	17.18554615	3.41E-66*
medianhousevalue	6.470E-07	6.91309E-08	9.359591006	8.00E-21*
pct_asian	-2.354E+00	0.275703715	-8.539371519	1.35E-17*
pct_blk	2.307E+00	0.035017136	65.89026857	<1E-300*
pct_native	1.869E+00	0.084238762	22.1833435	4.97E-109*
pct_owner_occ	1.026E+00	0.066403312	15.45180295	7.34E-54*
popdensity	5.527E-05	6.66216E-06	8.29555329	1.08E-16*
poverty	1.338E-01	0.093909634	1.424568375	1.54E-01
q_popdensity	6.223E-02	0.005668751	10.97691946	4.93E-28*
SVI_1_Socioeconomic	7.792E-04	0.012018797	0.064831177	9.48E-01
SVI_2_Household	-1.597E-01	0.022106948	-7.22532312	5.00E-13*
SVI_4_Housing	6.131E-01	0.019037185	32.20335838	1.58E-227*
SVI_AreaSqMiles	-3.855E-07	3.53522E-06	-0.10904808	9.13E-01
SVI_PctGE65	-2.379E-02	0.001689915	-14.07544728	5.38E-45*
SVI_PctLE17	3.921E-02	0.002051978	19.10783613	2.17E-81*
TRAVEL_DaytimePopDensity	-1.372E-05	4.13624E-06	-3.31602576	9.13E-04*
TwoWeekLag	2.719E-03	6.48201E-05	41.95433249	<1E-300

Supp. Table 7. Coefficient table for negative binomial model with daily deaths/(population size) as response, up to 6/11/2020, using both fixed and dynamic predictors, and standard errors evaluated by bootstrapping. *P<0.05 after Bonferroni correction.

Coefficients	**Estimate**	**Bootstrap SE**	**Z**	**P-value**
(Intercept)	-2.384E+01	0.511853521	-46.57473524	<1E-300*
AvgSocialDist	-1.980E-01	0.019271058	-10.27417594	9.21E-25*
AvgStatePctTested	2.578E+01	0.956732267	26.94492914	6.54E-160*
CHR_Diabetes	-5.614E-01	0.349540647	-1.60621703	1.08E-01
CHR_PrematureDeath	2.766E-05	7.18117E-06	3.852316808	1.17E-04*
CHR_Smoking	1.002E+00	0.455436756	2.199098851	2.79E-02
CHR_Traffic	-1.405E-05	2.55218E-05	-0.550641373	5.82E-01
education	2.774E-01	0.141983956	1.953637412	5.07E-02
hispanic	1.545E-01	0.110143785	1.402923989	1.61E-01
INSURANCE_PctTot	-2.082E-02	0.002767578	-7.524299453	5.30E-14*
log(SVI_TotPop)				
mean_pm25	9.807E-02	0.005269807	18.61055769	2.64E-77*
medhouseholdincome	3.054E-05	1.16622E-06	26.18960321	3.49E-151*
medianhousevalue	-1.382E-06	1.38846E-07	-9.956816096	2.35E-23*
pct_asian	-3.407E+00	0.393104539	-8.66652909	4.45E-18*
pct_blk	2.847E+00	0.073062197	38.96455145	<1E-300*
pct_native	3.283E+00	0.16844201	19.49062029	1.32E-84*
pct_owner_occ	2.576E+00	0.150208655	17.1520718	6.07E-66*
popdensity	1.418E-04	8.98854E-06	15.78022702	4.26E-56*
poverty	2.648E-01	0.243752888	1.086233377	2.77E-01
q_popdensity	1.676E-01	0.011942481	14.03586341	9.40E-45*
SVI_1_Socioeconomic	2.123E-03	0.030392755	0.069851213	9.44E-01
SVI_2_Household	2.205E-01	0.052725187	4.18180717	2.89E-05*
SVI_4_Housing	5.614E-01	0.040539776	13.84908105	1.29E-43*
SVI_AreaSqMiles	1.317E-05	6.05203E-06	2.175835248	2.96E-02
SVI_PctGE65	3.683E-02	0.003578858	10.28962054	7.85E-25*
SVI_PctLE17	1.820E-02	0.004967544	3.663370788	2.49E-04*
TRAVEL_DaytimePopDensity	-4.143E-05	5.84091E-06	-7.092374367	1.32E-12*
TwoWeekLag	1.901E-03	7.0937E-05	26.79788937	3.42E-158*

References

1. Lue, E. and J.P. Wilson, *Mapping fires and American Red Cross aid using demographic indicators of vulnerability.* Disasters, 2017. **41**(2): p. 409-426.
2. Horney, J., et al., *Assessing the Quality of Rural Hazard Mitigation Plans in the Southeastern United States.* Journal of Planning Education and Research, 2017. **37**(1): p. 56-65.
3. Sattar, N., I.B. McInnes, and J.J.V. McMurray, *Obesity Is a Risk Factor for Severe COVID-19 Infection: Multiple Potential Mechanisms.* Circulation, 2020. **142**(1): p. 4-6.
4. Wu, J., et al., *Quantifying the role of social distancing, personal protection and case detection in mitigating COVID-19 outbreak in Ontario, Canada.* J Math Ind, 2020. **10**(1): p. 15.
5. Project, T.C.T. 2020; Available from: https://covidtracking.com.
6. Unacast. *Social Distancing Scoreboard*. 2020; Available from: https://www.unacast.com/covid19/social-distancing-scoreboard.
7. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Infect Dis, 2020. **20**(5): p. 533-534.
8. Reif, D.M., et al., *Endocrine profiling and prioritization of environmental chemicals using ToxCast data.* Environ Health Perspect, 2010. **118**(12): p. 1714-20.
9. National Academies of Sciences, E., and Medicine, *Using 21st century science to improve risk-related evaluations.* National Academies Press, 2017.
10. Loomis, D., et al., *Carcinogenicity of lindane, DDT, and 2,4-dichlorophenoxyacetic acid.* Lancet Oncol, 2015. **16**(8): p. 891-2.
11. Marvel, S.W., et al., *ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models.* BMC Bioinformatics, 2018. **19**(1): p. 80.
12. Esri. *ArcGIS Javascript API version 4.13*. 2020.
13. Control, C.f.D.; Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/reopen-guidance.html.
14. Team, R.C., *R: A language and environment for statistical computing*. 2019, R Foundation for Statistical Computing: Vienna, Austria.
15. Wu, X., et al., *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv, 2020.
16. Williams, C.K.R., C.E., *Gaussian processes for machine learning* Vol. 2. 2006: MIT Press.
17. De Boor, C., *A practical guide to splines*. Applied Mathematical Sciences. 2001.
18. Wheeler, M.W., *Bayesian additive adaptive basis tensor product models for modeling high dimensional surfaces: an application to high-throughput toxicity testing.* Biometrics, 2019. **75**(1): p. 193-201.
19. Lang, S. and A. Brezger, *Bayesian P-splines.* Journal of Computational and Graphical Statistics, 2004. **13**(1): p. 183-212.
20. Moran, K.R. and Wheeler, M.W., *Fast increased fidelity approximate Gibbs samplers for Bayesian Gaussian process regression.* arXiv, 2020. **preprint: 2006.06537**
21. Ambikasaran, S. and E. Darve, *An Fast Direct Solver for Partial Hierarchically Semi-Separable Matrices.* Journal of Scientific Computing, 2013. **57**(3): p. 477-501.

22. Borm, S., *Efficient numerical methods for non-local operators: H2-matrix compression, algorithms and analysis* European Mathematical Society, 2010.
23. Quinn, S.C. and S. Kumar, *Health inequalities and infectious disease epidemics: a challenge for global health security.* Biosecur Bioterror, 2014. **12**(5): p. 263-73.
24. Verity, R., et al., *Estimates of the severity of coronavirus disease 2019: a model-based analysis.* Lancet Infect Dis, 2020. **20**(6): p. 669-677.
25. Coughlin, S.S., et al., *COVID-19 Among African Americans: From Preliminary Epidemiological Surveillance Data to Public Health Action.* Am J Public Health, 2020. **110**(8): p. 1157-1159.
26. Chowkwanyun, M. and A.L. Reed, Jr., *Racial Health Disparities and Covid-19 - Caution and Context.* N Engl J Med, 2020.
27. Centers for Disease Control, *Implementation of Mitigation Strategies for Communities with Local COVID-19 Transmission*. 2020; Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/community-mitigation.html.