CODE

BRIEFING PAPER

Using SDOH Data to Fight COVID-19 and Support Recovery Efforts

August 2020

Table of Contents

INTRODUCTION

The COVID-19 crisis has led to an explosion of data analysis and data-driven debate, perhaps more than any other event in recent history. Government agencies, nonprofits, and news organizations have collected both public health and social data, published it widely, and conducted myriad analyses to help guide decision-making about the pandemic. While these models have been valuable, they do not fully reflect a critical factor: the impact of social and economic context on the likelihood of contracting COVID-19, the availability of testing, and access to care for the disease.

A growing body of evidence shows that the social determinants of health have a major influence on an individual's health status — perhaps as much as, or even more than, the standard epidemiological measures that have been the basis of most COVID-19 models. These factors can help analyze the different risks COVID-19 or its mitigation pose to different communities, especially African American and Latino communities that already face poorer economic and health outcomes.

Using SDOH data to address COVID-19 can benefit the public in several ways. Epidemiologists and public health researchers can use the data to determine how social and environmental factors affect the spread and severity of the disease, and assess the health and societal consequences of social interventions for pandemic control. State and local public health departments and healthcare administrators can use it to better plan for the usage of hospital resources, or work to develop responsive economic and health models tailored to their communities. Policymakers will be able to use the data to better target interventions, such as through the identification of social characteristics that serve as risk or protective factors, and more effectively allocate limited resources to individuals and communities.

State health departments, healthcare companies, and academics are beginning to use SDOH data to predict COVID-19 risk in the populations they serve, at both a community and an individual level. However, organizations like these face a dual challenge: They need better data on the social determinants of health, and they need better data on the individuals who contract COVID-19 or who may be affected by the social consequences of the mitigation.

This Briefing Paper provided context for the *Roundtable on Using SDOH Data to Fight COVID-19 and Support Recovery Efforts*, and is also intended as a resource for anyone interested in this increasingly important topic. The Roundtable, held in August 2020, was co-hosted by the nonprofit Center for Open Data Enterprise (CODE) and the Office of the Chief Technology Officer (CTO) in the U.S. Department of Health and Human Services (HHS). It was designed to identify rapid opportunities to use social factors to predict and address COVID-19 at a local level. This Briefing Paper covers several key questions for the Roundtable and additional work:

- What is the state of data on COVID-19 and how can we improve it?
- What kinds of SDOH data do we need most urgently to fight COVID-19?
- How can SDOH insights help communities at high risk for COVID-19?

- How can we use data to help the U.S. healthcare system survive and recover from the pandemic?
- What state and local programs are providing models for fighting COVID-19?
- What can we learn from population-wide models and what are their limitations?
- Overall, how can we improve the data and models we use to fight the pandemic?

THE COVID-19 DATA LANDSCAPE

What is the state of data on COVID-19 and how can we improve it?

Case-based reporting helps to determine case numbers estimate infection rates, and track trends over time, geographic distributions, and outbreaks. COVID-19 case data is gathered at the individual, nursing home facility, hospital, and county/community levels. The National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC), and other stakeholders inside and outside HHS have developed new tools and systems to better track COVID case data. The recent transition of data collection responsibilities from the CDC may have impacts that remain to be seen. Regardless of who manages the data, however, there are a number of gaps and limitations that need to be addressed to meet the unique challenges of COVID-19 data collection.

Much of the country's current public health infrastructure was built to handle diseases like tuberculosis and measles, and therefore is not equipped to support a large-scale outbreak resulting from the emergence of a novel pathogen COVID-19. These outbreaks typically affected small, narrowly defined populations, orders of magnitude fewer than the millions of people affected in the case of this pandemic. The flaws in our current data collection systems[1] are evidenced by the difficulty in obtaining timely and accurate information on the number of infections, hospitalizations, and deaths. While medical records have undergone a transformation into electronic health records (EHR), lack of interoperability and other challenges make it difficult to use this data to track COVID-19.

Public health surveillance systems have generally suffered from underinvestment, and both national and state systems have failed to meet the challenge of tracking COVID-19. Nationally, the CDC has collected data on COVID-19 cases as an initial basis for tracking and analyzing the spread of COVID-19, but there have been gaps and inconsistencies in this data and required reporting. These include missing or incomplete information on disease severity and treatment, such as hospitalization status and ICU admission, as well as an absence of important patient characteristics, including underlying health conditions, age, race, gender, and ethnicity. At the state level, there is wide variation in the comprehensiveness of data collected on COVID-19 diagnoses, with only twelve states reporting comorbidities between deaths and eleven states reporting comorbidities by interacting two or more variables like gender, race, and age.

Some researchers are concerned that the current public health surveillance infrastructure could even collapse as the volume of COVID cases continue to rise. The Belfer Center has noted that a unified data infrastructure needs to be developed to support the reporting of COVID-19 cases from both testing venues and traditional medical centers. To be successful, this will need to have complete interoperability with the electronic medical systems of major hospital networks, support mobile surveillance testing, and be interoperable with other states' systems. This vision, however, cannot be realized in time to address the COVID-19 pandemic.

As a result of these gaps in data from conventional medical and public health sources, many groups including journalists, academics, private-sector groups, and citizen-scientists have taken the initiative to create their

own COVID-19 tracking systems. While these cannot fully substitute for public health data, they can provide some insights into the disease itself, its spread, and its impact on individuals with different risk factors. The table below shows some of the most widely used government and non-government sources for data on COVID-19.

Sources tracking COVID-19 cases in the US	
HHS Protect Public Data Hub[2]	Developed by the U.S. Department of Health and Human Services HHS Protect Project. The resource provides national hospital utilization, including hospital bed occupancy by state. The data is at the national and state level, and provides state FIPS code and HHS region.
CDC COVID Data Tracker[3]	This resource provides data on total cases and deaths, based on aggregate counts of COVID-19 cases reported by the states, the District of Columbia, New York City, and other U.S.-affiliated jurisdictions to the CDC.
World Health Organization (WHO) Coronavirus Disease (COVID-19) Dashboard[4]	Developed by WHO to track COVID-19 confirmed cases and deaths around the world.
John Hopkins University School of Medicine Coronavirus Resource Center (CRC)[5] Interactive Map	The CRC is a continuously updated source of COVID-19 data and expert guidance. They aggregate and analyze the best data available on COVID-19—including cases, as well as testing, contact tracing and vaccine efforts—to help the public, policymakers and healthcare professionals worldwide respond to the pandemic.
Data Repository for the 2019 Novel Coronavirus Cases[6]	Developed by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). In addition to county level time-series data, it reflects testing and hospitalization rates on a state level, incidence and mortality rates on a global level, and US state case curves.
Coronavirus Pandemic (COVID-19) Case Data[7]	Developed by Our World in Data. It is updated daily and includes data on confirmed cases and deaths from European Centre for Disease Prevention and Control (ECDC)[8], and testing from official reports and the COVID Tracking Project, other variables of potential interest such as age, poverty, GDP, etc.

Sources tracking COVID-19 cases in the US	
The COVID Tracking Project at the Atlantic[9]	The Atlantic is tracking racial and ethnic data from every state that reports it—and pushing those that don't to start. In collaboration with the BU Center for Antiracist Research[10], they analyzing this data to uncover the true impact of the outbreak on vulnerable communities.
New York Times[11] Interactive Map and Case Count	The New York Times is engaged in a comprehensive effort to track details about every reported case, deaths, and now probable deaths in the United States, collecting information from federal, state and local officials daily.
A County-level Dataset for Informing the United States' Response to COVID-19[12]	Developed by a research team from Cornell. It provides a machine-readable dataset that aggregates case data from governmental, journalistic, and academic sources on the county level, and contains over 300 variables that summarize population estimates, demographics, ethnicity, housing, education, employment and income, climate, transit scores, and healthcare system-related metrics.

While all these efforts are improving COVID-19 data collection, they face a basic limitation: the inadequacy of current COVID-19 testing. COVID-19 testing[13] programs have been slow to develop and still have much room for improvement. Since the beginning of the outbreak, the U.S. has failed to meet daily demands for testing, due to shortages of reagents, materials, and personal protection equipment (PPE). A recent analysis[14] has also shown that testing is particularly hard to access in minority communities, which are at greatest risk from the pandemic.

Testing data is essential in managing all aspects of this pandemic. For instance, this data serves as the foundation for most COVID-19 forecasting models, which predict future case surges and demand for emergency room services, hospital beds, ventilator equipment, and other forms of care. Without adequate testing data, forecasters are forced to rely on flawed data and their own assumptions.

NEW INSIGHTS FROM THE SOCIAL DETERMINANTS OF HEALTH

What kinds of SDOH data do we need most urgently to fight COVID-19?

The COVID-19 pandemic has drawn attention to the health disparities faced by low-income and minority communities, and these are largely caused by social, community, and environmental factors. The Social Determinants of Health[15] (SDOH) are defined as the "conditions in which people are born, grow, live, work and age that shape health." Research on the SDOH is a growing area of focus in the healthcare industry, since it holds the promise of both improving healthcare for underserved populations and helping to control cost.

The SDOH can determine one's day to day health and provide a more holistic understanding of factors that affect an individual's risk of disease and response to treatment. They can include measures as diverse as the air quality of a patient's neighborhood or their proximity to a grocery store. The economic and educational opportunities that people have access to often determine many of their health outcomes. Studying SDOH factors makes it possible to target at-risk communities to improve health and reduce healthcare costs.

The use of SDOH data has the ability to catalyze predictive analytics through improved location-based data about at-risk communities. This data can be collected directly from a patient in a clinical setting, and can be used in combination with an individual's EHR to better understand the possible risks they face. It can also be collected at the population level from a wide range of sources including federal, state, and local government agencies. Population-level SDOH data can be leveraged to develop an understanding of risks shared by groups of individuals in the same community or who share other characteristics.

There are many gaps in the availability of good SDOH data, as CODE described in the report from its October 2019 Roundtable on Leveraging Data on the Social Determinants of Health[16], which was co-hosted by HHS. That report identified the need to better define and standardize SDOH data, including the use of open source assessment tools and improved data governance; to create a sustainable infrastructure for SDOH data, including the involvement of organizations (CBOs); and to support local and state-based decision-makers using SDOH data.

There is also a growing need for better SDOH data, as well as COVID-19 data, at the state and local level. Some states, localities, health care systems, and CBO networks that have previously invested in SDOH infrastructure were able to quickly leverage that infrastructure to respond to the COVID-19 crisis. In many cases, state and local governments continue to improve existing infrastructure and rapidly prototype COVID tracking systems that uniquely draw on SDOH factors to understand how COVID 19 impacts their communities.

The table below from the Kaiser Family Foundation[17] shows the major categories of SDOH that are relevant to public health challenges.

Figure 1

Social Determinants of Health

Economic Stability	Neighborhood and Physical Environment	Education	Food	Community and Social Context	Health Care System
Employment	Housing	Literacy	Hunger	Social integration	Health coverage
Income	Transportation	Language	Access to healthy options	Support systems	Provider availability
Expenses	Safety	Early childhood education		Community engagement	Provider linguistic and cultural competency
Debt	Parks	Vocational training		Discrimination	
Medical bills	Playgrounds			Stress	Quality of care
Support	Walkability	Higher education			
	Zip code / geography				

Health Outcomes
Mortality, Morbidity, Life Expectancy, Health Care Expenditures, Health Status, Functional Limitations

KFF
HENRY J KAISER FAMILY FOUNDATION

The U.S. government and other public sources can provide much of the SDOH data that is relevant to COVID-19. The U.S. Census Bureau, for example, collects and publishes a wealth of data about America's changing population, housing, and workforce. The American Community Survey (ACS)[18] collects data on a number of SDOH factors at the ZIP and census-tract levels, making it usable for localized analysis. Some of these sources for SDOH data are shown in the table below.

SDOH Category	Potential linkages to COVID-19	Sample data sources
Economic Stability	Most workers receive health insurance through their jobs, but due to increasing rates of unemployment, people are losing their coverage. Only 11% of the employed lack health insurance, compared to roughly 30% of unemployed individuals.[19] Positive rates of COVID-19 and risk of severe infection are increased among low-income individuals as well.[20]	■ Department of Labor — Unemployment Statistics[21] ■ Internal Revenue Service (IRS) — Income Tax Statistics[22]
Neighborhood and Physical Environment	Homelessness[23] falls under this category and can be a predicting factor of COVID-19 infection. People who are homeless or have unstable living situations have a harder time partaking in social distancing, increasing their risk of contracting the virus. Additionally, communities with higher rates of homelessness, inadequate food access, poor environments, and other negative risk factors are predominantly black or minority populated.	■ Zillow — Walkability[24] and Housing data[25] ■ Department of Transportation — Proximity to Public Transportation[26] ■ National Parks Service data — Parks and Recreational Areas[27]
Education	The health and economic consequences of the pandemic for those with only a high school education or some college are much higher than for those with a bachelor's degree or higher.[28] The economic stress of unemployment can increase an individual's overall risk of illness. Low health literacy can also impact one's ability to access and utilize the latest scientific information and guidance on COVID-19.	■ National Center for Education Statistics — Dropout/graduation Rates[29] ■ American Community Survey — Educational Attainment[30]

SDOH Category	Potential linkages to COVID-19	Sample data sources
Food	Due to the pandemic, food supply chains have been affected at global, national, and local levels. This in turn is raising the number of people facing food insecurity in the U.S., which is a general risk factor for illness. There is also an increasing number of Americans on the Supplemental Nutrition Assistance Programs due to the pandemic.[31]	■ U.S. Department of Agriculture — WIC and SNAP enrollment rates by neighborhood[32] ■ American Community Survey — Public Assistance data[33]
Climate and Environment	There is a link between air pollution exposure and higher COVID-19 cases and deaths.[34] Researchers have shown that long-term exposure to pollutants, including nitrogen dioxide and sulphur dioxide, can reduce the functionality of lungs and increase susceptibility to respiratory illness. Such pollutants have shown to increase the risk of infection by viruses including COVID-19. Long-term exposure or childhood exposure to air pollution and asthma is also a risk factor for more severe and fatal illness.[35]	■ U.S. Environmental Protection Agency (EPA) — Air Quality and Toxic Waste[36] data ■ U.S. Department of Energy (DOE) data — Energy and Power Plants[37]

These and other data sources can help fuel new analyses of the link between SDOH and COVID-19. A number of companies in the healthcare space are now using COVID-19 case data together with SDOH data to predict the risk that certain individuals will be severely affected by COVID-19.

At least two population health management companies, Jvion[38] and ZeOmega[39], are using their own proprietary data together with public SDOH data to identify high-risk individuals within their patient populations. Because ZeOmega has access to insurance claims data and sometimes clinical data on tens of millions of individuals, they can analyze that data together with SDOH data to develop predictive risk models and iterate those models over time, for a number of medical conditions. ZeOmega has developed risk models for opioid overdose, diabetes onset and hospitalization risk, and is now applying the same approach to COVID-19. Similarly, Jvion, which describes itself as a predictive clinical AI company, combines clinical, claims, and SDOH data on tens of millions of people to understand their risk holistically across a multitude of clinical use cases. These clinical AI/machine learning models identify underlying risk factors and inform appropriate interventions. The models can leverage SDOH data and individual questionnaire responses to understand risk/vulnerability at a census tract level.

These models are localized by nature: They require either individual, questionnaire-based SDOH data or highly localized SDOH data, at a census tract level at least, in order to have sufficient granularity for analysis. Another company, Socially Determined[40], has developed analyses of SDOH data at a highly localized level for geographic locations across the U.S. They are now working with a number of health plans to apply that data for risk analysis of COVID-19 and other health conditions.

While the work being done by health information companies is still largely proprietary, a number of organizations are beginning to develop public models that use SDOH data to analyze COVID-19, as shown in the table below. Some of these models have been used for state and local interventions that have effectively incorporated SDOH data in programs that have sought to address the spread of COVID-19.

Model Name	Description	Partners	Resulting Interventions
COVID-Net[41]	Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET) is a population-based surveillance system. It collects data on laboratory-confirmed COVID-19-associated hospitalizations among children and adults through a network of over 250 acute-care hospitals in 14 states.	Emerging Infections Program (EIP)[42], the Influenza Hospitalization Surveillance Project (IHSP)[43]	COVID-Net has been used to develop risk factors[44] for intensive care unit admission and in-hospital mortality among adults. A study[45] has also been conducted using this data to identify hospitalization rates and characteristics of patients hospitalized due to COVID-19.
Surgo Foundation Community Vulnerability Index (CCVI)[46]	The CCVI builds CDC's Social Vulnerability Index (SVI), a validated metric to help policymakers and public health officials respond to disasters, including disease outbreaks. It incorporates the SVI's sociodemographic variables, along with risk factors specific to COVID-19 and variables measuring the capacity of public health systems.	CDC[47], Centers for Medicare & Medicaid Services[48] (CMS), the Harvard Global Health Institute[49], PolicyMap[50], the US Bureau of Labor Statistics[51] (BLS), the US Census Bureau[52] (USCB), and the Association of Public Health Laboratories[53]	The CCVI combined with the SVI's metric results in six core themes that together account for 34 factors that make a community vulnerable to the COVID-19 pandemic. The scores for the six themes and the overall CCVI are available here[54] by census tract, county, and state.

Model Name	Description	Partners	Resulting Interventions
JVION COVID Community Vulnerability Map[55]	The COVID Community Vulnerability Map is a publicly available interactive map that identifies pockets of individuals and communities across the U.S. at risk for experiencing severe outcomes ranging from hospitalization to mortality as a result of contracting a respiratory infection like COVID. The map also provides the socioeconomic factors influencing that risk.	Microsoft[56], Johns Hopkins Coronavirus Resource Center[57]	Insights from this map can inform providers, public health organizations and community support agencies as they look to deploy interventions, outreach and other services to keep individuals from contracting the virus and once infected manage towards a positive outcome
Rensselaer Polytechnic Institute (RPI) COVIDMINDER[58]	COVIDMINDER reveals the regional disparities in outcomes, determinants, and mediations of the COVID-19 pandemic. *Outcomes* are the direct effects of COVID-19. Social and *Economic Determinants* are pre-existing risk factors that impact COVID-19 outcomes. *Mediations* are resources and programs used to combat the pandemic. This RPI project measures racial disparities in New York and Connecticut and layers information related to disease factors such as diabetes, and healthcare system capacity measures such as hospital beds.	The Rensselaer Institute for Data Exploration and Applications[59] (IDEA), United Health Foundation[60]	By comparing how factors such as health behaviors, health-care access, and socioeconomic status contribute to the spread of COVID-19 using data visualizations, COVIDMINDER is able to reveal regional disparities in disease determinants, public health measures, and outcomes.

THE IMPACT ON LOW-INCOME AND MINORITY COMMUNITIES

The impact of COVID-19 on minority and low-income communities shows how profoundly social and economic factors are shaping this pandemic. But while we know that the social determinants of health (SDOH) are putting vulnerable groups at risk, we don't know which factors are responsible, how important those factors are, or how to address them to combat COVID-19. This section describes some of the key SDOH factors that may disproportionately put low income and minority communities at risk of contracting COVID-19.

How can SDOH insights help communities at high risk for COVID-19?

Racial and ethnic minorities are more likely than non-Hispanic whites to be poor or near poor. The CDC[61] has identified these groups to be at an increased risk of contracting COVID-19 due to living conditions, work circumstances, and health circumstances. These three core factors all fall under the jurisdiction of SDOH, and are necessary for an in-depth analysis of how systemic factors impact one's likelihood of contracting and dying from the virus. For example, the Latinx community is being hit harder by the virus: Official CDC statistics show that this community is being hospitalized and dying at four times the rate[62] of their non-Hispanic white counterparts. Another analysis by NPR found that 32 states, plus Washington D.C., reported that African Americans are also dying at higher rates[63] than their proportion of the population. For instance, as of May 2020, the state of Wisconsin reported 141 African American deaths from COVID-19, representing 27 percent of all in-state deaths, while Blacks only make up 6 percent of the state's population. The HHS Office of Minority Health (OMH) recently announced a $40 million award grant given to Morehouse University School of Medicine to fight COVID-19 in racial and ethnic minority groups, and rural and vulnerable communities.[64] Morehouse School of Medicine will be working with the OMH on the National Infrastructure for Mitigating the Impact of COVID-19 within Racial and Ethnic Minority Communities (NIMIC) Initiative.

Low income is also an independent risk factor for severe COVID-19 infection. A recent study by the Centers for Medicare and Medicaid Services[65] showed that people who qualified for Medicaid as well as Medicare - an indication of low income - were four times as likely to have been infected or hospitalized with the virus as those on Medicare alone. For minority and low-income communities especially, understanding SDOH factors can be critical to understanding the risk of COVID-19 and designing interventions to reduce the risk of infection and the severity of the illness.

Living Conditions. The living conditions of racial and ethnic minority groups can greatly contribute to increased health risk and make it more difficult to ensure safety guidelines for the prevention of COVID-19 and ability to seek care. Racial and ethnic minorities are more likely to live in densely populated areas[66], due to structural and institutional racism in the form of residential housing segregation. Overcrowding is also more likely in tribal reservation homes and Alaska Native villages. Certain racial and ethnic minority groups are also overrepresented in jails, prisons, homeless shelters, and detention centers, where they are forced to live, work, eat, and recreate within congregate spaces.

Racial housing segregation is linked to greater negative health outcomes, including asthma and other underlying medical conditions, due to pollution and other environmental factors in the locations of segregated communities. These health conditions place people at an increased risk of becoming severely ill or dying from COVID-19. Many racial and ethnic minority groups additionally live in neighborhoods that are in food deserts[67] and are located far from medical facilities. These neighborhoods may also lack safe and reliable transportation, making it more difficult to stockpile supplies that would allow people living there to stay home and receive care if sick.

Work Circumstances. Certain jobs and workplace policies place workers at an increased risk of contracting COVID-19, such as essential worker positions. A higher proportion of members of some racial and ethnic minority groups are more likely to hold these positions. For example, the risk of infection may be greater for workers in essential industries and businesses, including healthcare, grocery stores, and factories. These workers must still show up to work sites despite outbreaks in their communities, and some have no choice but to continue working in these jobs due to their economic circumstances. Workers who do not have paid sick leave, like many essential workers, are more likely to keep working despite being sick. On average, racial and ethnic minorities earn less than non-Hispanic whites, have less accumulated wealth, lower levels of educational attainment, and higher rates of unemployment. These factors can each play a role in the quality of the social and physical conditions in which people live, learn, work, and play, and can have an impact on health outcomes.

Health Circumstances. Health and healthcare inequities disproportionately affect racial and ethnic minority groups. Some of these inequities can put people at increased risk of becoming severely ill or dying from COVID-19. Hispanics are almost three times as likely to be uninsured[68], as non-Hispanic whites, and Black people are almost twice as likely to be uninsured. Similarly, Black people of all age groups are more likely than non-Hispanic whites to report not being able to see a doctor in the past year due to costs.

A consistent barrier these communities face is the reluctance to seek medical care because of distrust of the healthcare system, language barriers, or cost of missing work. Compared to non-Hispanic whites, Black people experience greater rates of chronic conditions and higher death rates at early ages. Similarly, American Indian and Alaska Native adults are more likely to be obese, have high blood pressure, and smoke cigarettes than non-Hispanic white adults. Racism, stigma, and systemic inequities undermine prevention efforts, increase levels of chronic and toxic stress, and ultimately sustain health and healthcare inequities.

A better understanding of SDOH factors can help design mitigation strategies, such as quarantine, social distancing, and temporary business closures, that take the circumstances of different communities into account. Many of the mitigation measures put in place by the government to curb COVID-19 were not well designed for low-income families and individuals. For example, it is more common for low-income minorities in the U.S. to live in multigenerational or multi-familial households[69]. These living conditions, combined with relatively small living spaces, make it more difficult to quarantine if an individual has or is suspected to have COVID-19. Social distancing is another measure that cannot be carried out as effectively in low-income minority communities. These communities have more essential workers, don't have as much access to private transportation, and utilize public transportation to a higher degree. This in turn makes it much more difficult to adequately keep a social distance[70] of six feet, as recommended by the CDC.

IMPROVING HEALTHCARE SYSTEM RESILIENCE

How can we use data to help healthcare systems survive and recover from the pandemic?

The COVID-19 Pandemic has proven to be a stress test that has stretched the boundaries of the U.S. Healthcare system and revealed large gaps in both data collection and healthcare delivery. This shock to the healthcare system has sparked renewed conversations about the value of healthcare resilience. From the Well Being Trust's recently launched Thriving Together[71] initiative, which acts as a springboard for healthcare equity and resilience, to renewed conversations in the private sector about planning, healthcare survival and recovery are fresh in the imaginations of many. This section reviews some of the literature around healthcare resilience and how SDOH factors may play a role in better understanding this emerging field.

COVID-19 is part of a variety of external challenges such as antimicrobial resistance, financial burdens, extreme climate events, and larger disease outbreaks that have put pressure on the healthcare system. The term "health system resilience" has been used to understand the inability of health systems across the United States and in global settings to respond to COVID-19 surges or effectively share information that would help combat the epidemic. For example, the United States Agency for International Development (USAID) aims to promote health resilience in international settings by being able to distribute commodities to areas in need, redeploy key human and financial resources for vulnerable populations, and work across government sectors to engage community leaders and other key stakeholders.[72]

The Office of the Assistant Secretary for Health is currently seeking information[73] to help better understand how key stakeholders have defined resilience through "their use of data, analytic approaches, and proven indicators." MIT professor David Simchi-Levi has noted that like other industries, health system resilience may be reflected in two major metrics[74]:
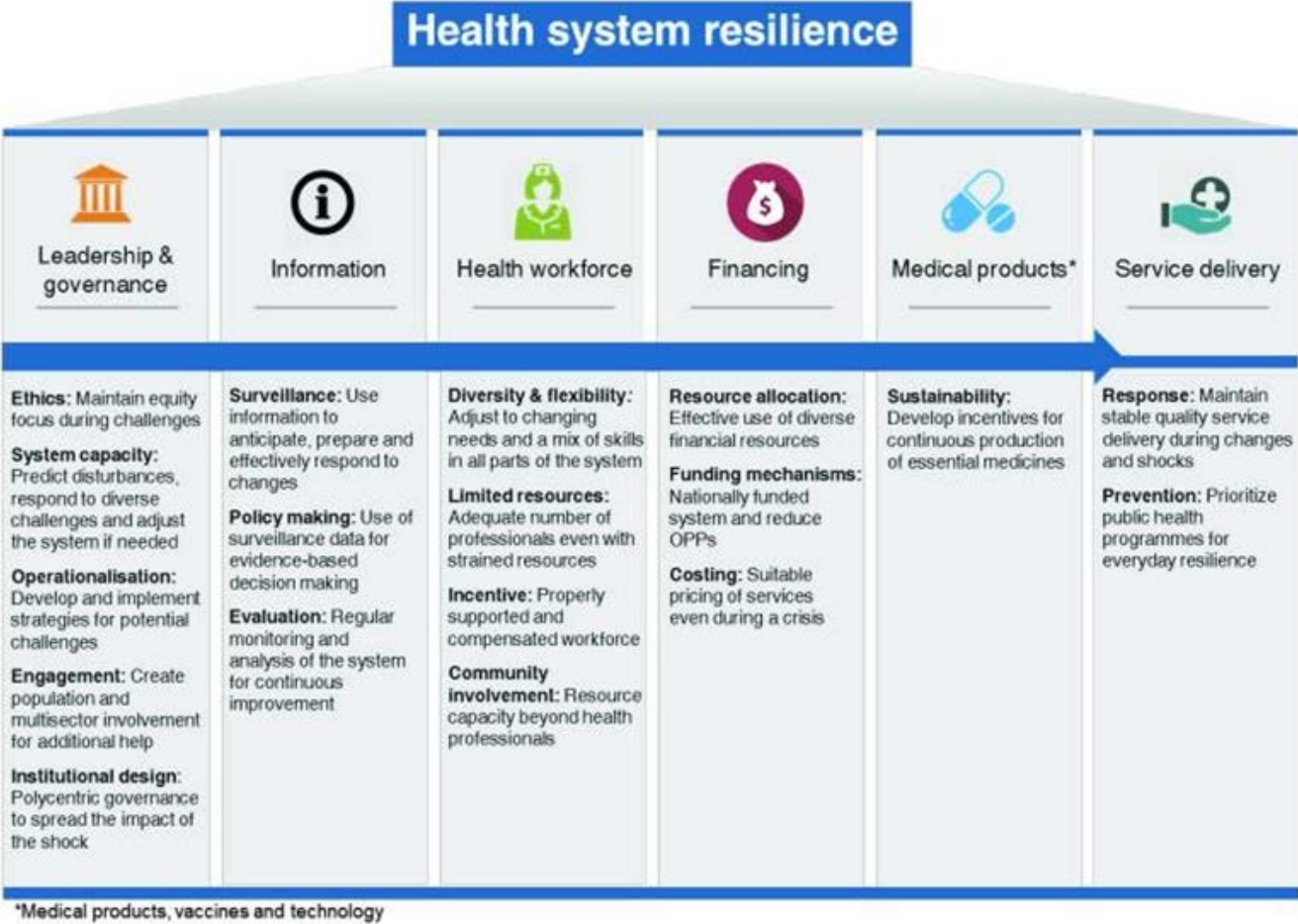
- **The Time to Survive:** This metric seeks to better understand how long an enterprise can survive when there is a shortage of some critical good. How long can a clinic, hospital, or healthcare network survive without access to ICU beds, PPE, or ventilators when intaking patients?

- **The Time to Recover:** This metric seeks to understand how long it will take for a system to properly restore this shortage of some critical good.

The concept of health system resilience indicates the ability of a health system to respond to extreme changes or shocks without the possibility of collapse or lack of function. A paper from the NIH[75] points out that the overall definition of resilience is "the capacity of an individual, population or system to absorb a shock, while still retaining the fundamental functions or characteristics of the original state." This definition, however, has been critiqued for not incorporating possible changes in capacity or the ability to adapt to a new state. Other definitions of resilience, especially for healthcare, have claimed that resilience should incorporate adaptive and transformative capabilities that adjust capacity to anticipate future shocks.

Authors of the NIH paper have adopted and defined key resilience metrics from the World Health

Organization's Framework that summarize the health systems six major functions: leadership and governance, information, health workforce, financing, medical products, and service delivery. Practitioners have sometimes approached these six attributes and incorporated specific dimensions and metrics for these six functions, as the figure below displays.

Health system resilience

Leadership & governance	Information	Health workforce	Financing	Medical products*	Service delivery
Ethics: Maintain equity focus during challenges **System capacity:** Predict disturbances, respond to diverse challenges and adjust the system if needed **Operationalisation:** Develop and implement strategies for potential challenges **Engagement:** Create population and multisector involvement for additional help **Institutional design:** Polycentric governance to spread the impact of the shock	**Surveillance:** Use information to anticipate, prepare and effectively respond to changes **Policy making:** Use of surveillance data for evidence-based decision making **Evaluation:** Regular monitoring and analysis of the system for continuous improvement	**Diversity & flexibility:** Adjust to changing needs and a mix of skills in all parts of the system **Limited resources:** Adequate number of professionals even with strained resources **Incentive:** Properly supported and compensated workforce **Community involvement:** Resource capacity beyond health professionals	**Resource allocation:** Effective use of diverse financial resources **Funding mechanisms:** Nationally funded system and reduce OPPs **Costing:** Suitable pricing of services even during a crisis	**Sustainability:** Develop incentives for continuous production of essential medicines	**Response:** Maintain stable quality service delivery during changes and shocks **Prevention:** Prioritize public health programmes for everyday resilience

*Medical products, vaccines and technology

Information may be one of the most important attributes of healthcare resilience. A variety of studies that have sought to understand healthcare resilience have emphasized the need for timely surveillance data in order to enact relevant and effective mitigation measures and policies. Moreover, the WHO model highlights service delivery as a critical factor that is dependent on the other five functions.

Researchers have attempted to develop models for some time to better understand how healthcare systems are able to respond to major crises. In the current COVID-19 pandemic, researchers and policymakers must consider a wide array of measures from measuring personnel and hospital staff to the volume of equipment a hospital may have to respond to the crisis.

Some Researchers have sought to understand hospital capacity and demand to model resilience. For example, a model developed by researchers at Colorado State University[76] aimed to predict resilience in the event of an earthquake. Their model incorporated a number of key factors such as the number

of staffed beds, hospital staff availability, housing functionality, patient waiting time for treatment, and also the probability of patient X going to healthcare facility Y. They also tried to factor in some of the environmental and physical conditions that could impact a healthcare system from electric power to the strength of their telecommunications system.

In the private sector, Facebook AI[77] has partnered with New York University's Courant Institute of Mathematical Sciences to create localized forecasting models of the spread of COVID-19. The researchers used testing data published by the State of New Jersey[78] and State of New York[79], and applied sophisticated analytic models to account for relationships among counties. To build hospital-level COVID-19 forecasts for medical resource allocation, Facebook is also collaborating with NYU Langone Health's Predictive Analytics Unit and Department of Radiology to develop models that can learn from de-identified clinical data, and then share open-source predictive algorithms so that hospitals can train models on their own data. Facebook's models are helpful since they make local predictions on a county and hospital level. The detailed AI algorithms have not yet been made public, as the research team continues evaluating other sources of data, such as Mobility Data Network Map[80] from Facebook's Data for Good team, to see whether they help improve the model's performance.

While these and other models to predict health system resilience are promising, models need to be fed with reliable data. Some data is in short supply currently, especially without widespread US testing for the novel coronavirus that causes COVID-19. For example, it is unknown how many people have been infected without symptoms. Other inputs, such as incubation periods and death rates, change by the day as we learn more about the virus. Human factors also make the modeling challenging. Individual behaviors, health care infrastructure and political response can all affect the outcome of an epidemic.

Apart from the basic medical components of healthcare systems, SDOH factors like transportation, access to food, and economic stability all impact healthcare system resilience. Models of healthcare resilience are showing that SDOH factors are critical to understanding how healthcare systems can survive and recover from pandemics. Researchers have noted that planners should incorporate key socioeconomic data into disease surveillance systems in order to measure how certain communities will be affected both by COVID-19 and by potential future pandemics and disease outbreaks. Some key factors include the following.

Transportation and National Infrastructure. Transportation ensures that HHS and other federal actors can rapidly distribute PPEs, vaccines, and other key equipment to hospitals and healthcare institutions around the country. Factoring in transportation systems also can pinpoint how different communities will respond to a shock like COVID-19, from urban transportation systems that might function as vectors of transmission to addressing rural communities that may have little access to transportation in the event they need to visit a healthcare facility.

Climate Data and the Built Environment. A recent report from the Natural Resources Defense Council[81] noted that climate data and planning can support healthcare resilience planning in two respects. First, climate scientists have long attempted to model how a variety of institutions and supply chains would respond to a sudden climate event such as a natural disaster. These models may provide guidance to healthcare planners. Secondly, there is a growing connection between climate events and how they

impact public health, from heat waves that could impact COVID-19 transmission to how air pollution has functioned as a potential comorbid factor for COVID-19. The CDC currently has a Climate-Ready States and Cities Initiative[82] that provides "public health expertise to help state and city health departments prepare for and respond to the health effects that a changing climate may bring to their communities."

Access to Food and Food Distribution. The food distribution system of the United States is an important part of the necessary infrastructure needed to ensure the ongoing health of communities. During the pandemic, it is also essential to distribute food to communities that are impacted by virus mitigation measures.

STATE AND LOCAL ACTION

What state and local programs are providing models for fighting COVID-19?

State health departments are developing models to predict COVID-19 risk in the populations they serve, at both a community and an individual level. There is a growing need to rapidly enable state and local governments to expand their data collection efforts and coordinate these projects to provide more accessible data for widespread analysis. Current projects in Louisiana, Indiana, and California provide models that may be useful to other states as well.

Kentucky Coronavirus Monitoring and State Model

This resource was developed by the Team Kentucky Fund in partnership with the Kentucky Department for Public Health and Kentucky Cabinet for Health and Family Services.[83] Its purpose is to provide COVID-monitoring data, resources like financial assistance applications and testing appointments, and information on the disease. The model provides both probable and lab-confirmed positive numbers of cases, deaths, and total tested, with an interactive map available by county as well. Kentucky residents can also look up labs that provide COVID-19 testing and services by county. The source collects updated numbers from around the state by allowing healthcare facilities to submit updates through their site. Team Kentucky also provides information on telehealth services, contact tracing, and other resources to help guide residents through the health and economic crisis.[84]

Louisiana Health Equity Task Force[85]

The governor of the State of Louisiana announced the launch of their COVID-19 Health Equity Task Force this past April, in collaboration with universities, research institutions, and medical experts. The program was developed in response to the disproportionate COVID-19-related deaths among the state, and vulnerable communities. The purpose of the task force is to provide recommendations relative to health inequities which are affecting communities that are most impacted by the coronavirus. The aim is to develop interventions which provide greater access to high quality medical care and improve health outcomes. The desired outcomes of this program are to provide reliable and data driven information on COVID-19 safety and prevention, provide the medical community with best practices and protocols for treating communities with underlying medical conditions and health disparities, and ensure testing availability and ease of access for all communities. The taskforce is beginning its work but has not yet published data.

Indiana Virtual Care at Home Model[86]

Indiana's Family and Social Services Administration developed a Virtual Care at Home model connecting individuals with COVID to federally qualified health centers for monitoring and virtual care. This allows for better hospital surge planning, and allows providers to care for patients without them having to leave their homes. The pilot, conducted in partnership with community health centers and federally qualified health

centers within Indiana, has gone live in three regions. Indiana's Governor, Eric Holcumb, also announced the integration of 'Indiana 2-1-1[87]' which is a free confidential service that connects residents with community services and resources. The service was launched in April 2020 and had served nearly 16,000 Indiana residents by July.

University of California, San Francisco (UCSF) Health Atlas[88]

The Health Atlas provides data on various domains of social determinants of health, relevant health outcomes, and COVID-19 case, death, and hospitalization data for the state of California. All reported data are publicly available. SDOH data includes demographic, socioeconomic, community, neighborhood, and healthcare data at the Census Tract level. By collecting COVID-19 data and incorporating the SDOH, the Health Atlas has been able to identify individuals and communities that are more at risk of contracting the disease. UCSF has also been able to measure the economic impact of COVID in California.

MODELING THE SPREAD OF THE DISEASE

What can we learn from population-wide models and what are their limitations?

This Briefing Paper has described the need for more targeted and sophisticated analyses of data related to COVID-19, including the need for analyses that incorporate SDOH data, that focus on minority and low-income communities, and that help predict and improve health system resilience. All of that work will take place in the larger context of population-wide models that are now tracking the disease and predicting its spread.

Descriptive and predictive models can be helpful to make informed guesses about the COVID-19 disease, its future spread, and effects of different actions and interventions to plan ahead with effective decision-making. Given the lack of proven effective drug treatments for COVID-19, many models examine effects of non-pharmaceutical interventions (NPIs) on human behavior. Researchers from academia and industry have utilized key epidemiological modeling approaches including:

- Curve-fitting extrapolation. The model infers epidemic trends in a given location by fitting the observed data, and then applies mathematical rules to construct curves to approximate the future epidemic path. The assumptions may be based on experiences in other locations, and/or local factors of populations, disease transmission, etc.

- Agent-based models. The model creates a synthetic community and simulates the interactions of individuals/collective entities ("agents") in that community. The agents are given traits and initial behavioral rules to organize their movements, mixing patterns, health interventions, etc.

- SEIR. The model divides an estimated population into different groups of "susceptible (S)", "exposed (E)" , "infected (I)" and "recovered (R)", and then applies mathematical rules about how people move from one compartment to another, using assumptions about the disease process, social mixing, public health policies, etc. Models that do not include estimates of exposed populations are "SIR" models.

Experts are also leveraging AI techniques to improve model accuracy. Each model has its own strengths and limitations. The table below summarizes several of the more widely cited models, which are described in detail in the Appendix.

Model and Organization Responsible	Primary Approach	Model Description	Pros/Cons
John Hopkins University School of Medicine Coronavirus Resource Center (CRC)[89] Interactive map	SEIR	■ Predicts cases, deaths, hospital and ICU admissions, bed use, ventilator use under different NPIs ■ Uses assumptions about disease incubation period, infectious period, fatality rate, hospitalized and ICU patients, and ICU admissions that are ventilated	■ Generic model, to be applied to various spatial scales given shapefiles, populations, cases ■ Lack of differentiated assumptions about fatality rate across age, gender, underlying illness and access to health care ■ Uncertainty due to a lack of timely testing data ■ Generalized assumptions about the strictness and effectiveness of different NPIs across regions
Institute for Health Metrics and Evaluation (IHME) COVID-19 Model[90]	Curve-fitting/extrapolation	■ Predicts number of hospitalizations and deaths in the U.S. by state for the next four months. ■ Uses daily confirmed COVID-19 deaths from WHO websites and local/national governments, hospital capacity and utilization by state, observed COVID-19 utilization data from select locations.	■ Assumes same social distancing policies across regions, when these may vary in actuality ■ Assumes stringent social distancing regulation ■ Large uncertainty bands due to inaccurate temporal data on mortality, hospitalization counts, etc.

Model and Organization Responsible	Primary Approach	Model Description	Pros/Cons
Los Alamos National Laboratory (LANL) COVID-19 Cases and Deaths Forecasting Model[91]	Curve-fitting/ extrapolation	■ Predicts cases and deaths by state ■ Uses assumptions about the presence and growth rate of social distancing interventions	■ More robust assessments based on collective statewide estimations ■ Unstable forecasts for states with under 100 confirmed cases and/or 10 confirmed deaths ■ Cannot model intervention effects for scenario analysis ■ Wide prediction intervals
Imperial College NPI Model[92]	SEIR	■ Predicts cases, deaths across different mitigation and suppression scenarios, over the next year ■ Uses assumptions about the disease process, social mixing, public health policies, etc.	■ Assumptions on the parameters match real-world observation ■ Imprecise measure of disease infectivity ■ Lack of reliable testing dataset of infected patients without showing symptoms

Model and Organization Responsible	Primary Approach	Model Description	Pros/Cons
Columbia University COVID-19 Risk[93] Model & Mapping Tool	SEIR	■ Projects the number of severe cases, hospitalizations, critical care, ICU use, and deaths under different social-distancing scenarios, for 3-week and 6-week periods ■ Uses county-to-county work commuting data from the US Census Bureau, SafeGraph estimates of the reduction of inter-county visitors.	■ Strong county-level projection ■ Dynamic simulation with 4 models with varying assumptions on reproduction ■ Unobserved policy effects within 2-week lag between infection acquisition and case confirmation ■ Lack of public data on staffing or ventilator supplies by states
COVID-19 Projection Model[94] by Northeastern University, Fogarty International Center, Fred Hutchison Cancer Center, University of Florida, etc.	Agent-based / Global Epidemic and Mobility Model (GLEAM)	■ Projects cases and deaths by state, under no-mitigation v.s. "stay-at-home" scenario ■ Estimates effects of school closures, smart working, social distancing on virus transmissibility ■ Uses sociodemographic data (e.g. public census data, survey results), to construct representative synthetic state populations	■ Does not include pre-symptomatic transmission ■ Estimates on "stay-at-home" impacts derived from Chinese contact data, which may not apply universally ■ Does not account for factors such as contact tracing, seasonal temperature/humidity, superspreading event, differential transmissibility across age brackets

TOWARDS BETTER INDICATORS AND MODELS: OPPORTUNITIES FOR IMPROVEMENT AND DISCUSSION

Overall, how can we improve the data and models we use to fight the pandemic?

The explosion of data that has resulted from COVID-19 provides an unprecedented opportunity to both improve existing models and identify major gaps in data collection that could provide needed information for policymakers, civil society, researchers, and the medical community. As part of this Briefing Paper and to prepare for the discussion for the August Roundtable, CODE has included some possible options to improve existing models and datasets. CODE has identified two overall, cross-cutting goals to begin:

Provide reliable testing datasets

- Provide up-to-date confirmed cases, pre-symptomatic cases, fatality rates, recovery rates, reinfection rates, reproduction rates, etc.
- Provide detailed measures and statistics by key demographic groups (age, race, gender, etc).
- Provide anonymous data from confirmed patients, including patient demographics, medical histories, and unstructured data like clinical notes, X-rays, CT-scans to fit AI-driven models
- Collect local data on a city/county/hospital level for accurate predictions with collective estimates

Incorporate potential SDOH risk factors into the model

- Aggregate socio-demographic datasets (age, gender, race, population density, GDP, household-income, education level, etc.) into a machine-readable format for convenient modeling.
- Concatenate different non-pharmaceutical intervention policies (e.g. social distancing, travel restriction, airport screening, etc.) with varying degrees of strictness across states/counties, better to specify the expected date or detailed threshold to lift the restriction
- Trace contact patterns based on mobility/transportation data for different purposes (e.g. working, education, shopping, etc) on an international/national/state/county level.
- Leverage social media data to examine people's location and anticipate their positive/negative sentiments towards the virus transmission
- Differentiate model assumptions across demographic brackets
- Provide public data on staffing or ventilator supplies by states/counties

In addition to generally improving the models themselves and incorporating new data, a number of groups have expressed the value of expanding the array of indicators that policymakers, civil society, and researchers should have access to. For example, Resolve to Save Lives, an initiative of Vital Strategies, has developed a comprehensive list of 15 indicators[95] disaggregated by race, ethnicity, income, sex, and ZIP codes. Sample indicators include:

- COVID-19 daily hospitalization per capita rates and 7-day moving average

- Percentage of new cases from among quarantined contacts, by week,

- New infections among health care workers not confirmed to have been contracted outside of the workplace, by week.

Additionally, HHS has issued guidance on what indicators[96] it should collect from the hospitals and other healthcare institutions.

The Harvard GenderSci Lab has developed a scoring scheme and Report Card to evaluate the comprehensiveness of socially relevant, intersectional data publicly reported by each state. The scoring scheme includes variables on age, sex/gender, race/ethnicity, and comorbidity status, as well as their interactions. Socioeconomic status, occupation, sexual orientation, and immigrations status were identified as crucial variables as well. The Harvard group was able to show that adding variables through intersectional analyses improved the ability to detect patterns in the disease[97] to allow for a more nuanced understanding of the diverse risk factors for COVID-19. Moreover, the GenderSci Lab observed that the inclusion of U.S. territories and data on Indian Reservations is essential[98] to assess the availability of socially relevant US surveillance data.

Possible Questions for Discussion

CODE is now developing a set of questions for discussion at the Roundtable around four core themes. The following questions are both a preview of the Roundtable, and suggested questions that all groups addressing the COVID-19 pandemic may want to examine.

Addressing the Impact on Low-Income and Minority Communities

- What policy levers at the federal, state, and local level can be used to help vulnerable persons financially and through social services? What approaches can be taken to overcome challenges in implementing such policies?

- What are the effective short-term strategies for local, state, and federal entities to improve how COVID-19 data is collected and reported using CDC and OMB standards for race/ethnicity categories? In addition to race/ethnicity, how can we support further stratification by income, geographic location, disability status, neighborhood characteristics?

Using Data for Healthcare System Resilience and Recovery

- What kinds of SDOH data should be prioritized for research on healthcare capacity, access, and resiliency – such as testing predictive models for how many people will suffer symptoms severe enough to require hospitalization, surge planning, and where they will risk overwhelming their local hospitals' capacity or avoid accessing care?

- How can we also identify healthcare providers with excess capacity that may be able to help more stressed areas?

Data-Driven Action at the State and Local Level

- What kinds of federal data could be useful to states in the short term, and what local, state and national-level surveillance data will be important for rapid learning related to effective response and public health and clinical interventions?

- What are the best opportunities for public-private collaboration in research and data-driven approaches to COVID-19 that leverage SDOH data?

Assessing and Improving Health Data Interoperability and Resources for SDOH and COVID

- What new online tools, such as websites or GitHub repositories, would help support and accelerate this work? What are some new/emerging non-proprietary based health IT tools that can be leveraged for further scale and dissemination?

- What barriers impede data sharing, in all dimensions (i.e. interoperability, HIPAA)?

- How can experts and stakeholders outside of government help CMS, ONC, and other HHS offices in their work to improve SDOH data? How can they help state and local stakeholders?

- How can federal and stakeholder efforts to advance SDOH data use and health IT interoperability support standardized approaches to COVID and SDOH data?

ACKNOWLEDGMENTS

CODE is an independent nonprofit organization based in Washington, D.C. whose mission is to maximize the value of open government data for the public good. CODE believes that open government data is a powerful tool for economic growth, social benefit, and scientific research. Since 2015 CODE has worked with the White House, numerous U.S. federal agencies, and several national governments and NGOs around the world to help them improve how they collect, publish, and apply data to better meet the needs of their data users. For more information, please visit **www.OpenDataEnterprise.org**.

APPENDIX: AN IN-DEPTH LOOK AT POPULATION-WIDE MODELS

The public understanding of COVID-19, and the public response to the pandemic, has been largely shaped by a number of population-wide models designed to track and predict the spread of the virus and its impact. This Appendix describes several of the most prominent models, their usefulness, and their current limitations.

The Johns Hopkins University (JHU) COVID-19 Scenario Pipeline Model

The JHU COVID-19 Scenario Pipeline Model[99] uses the *SEIR* method to project US cases, deaths, hospital and ICU admissions and bed use, ventilator use under different suites of Non-pharmaceutical intervention interventions. The model divides an estimated population into different groups of "susceptible (S)", "exposed (E)", "infected (I)" and "recovered (R)", and then applies mathematical rules about how people move from one compartment to another. General model assumptions include disease incubation period, infectious period, fatality rate, hospitalized and ICU patients, and ICU admissions that are ventilated.

The model is designed to be generic so that it can be applied to different spatial scales, given shapefiles, population data, and COVID-19 confirmed case data. (For example, it has been deployed in the state of California[100].) However, it relies on assumptions that may not capture the details of real-world observations. For instance, fatality rates are not one-size-fits-all as the model assumes; they differ by SDOH factors such as age, gender, underlying illness and access to health care. Without widespread and timely testing across different regions, there is considerable uncertainty in the model's predictions since it is unclear how the virus behaves, and how many people are infected or pre-symptomatic. The model also assumes that social distancing starts across a state all at once, while in fact, some counties within a state may order residents to shelter in place earlier than others. It also assumes that the effectiveness of social distancing measures in a given state decreases by roughly 25% after those orders are lifted. This estimate may not reflect all areas, as the effect depends on policy strictness and residents' behavioral patterns across regions.

Institute for Health Metrics and Evaluation (IHME) COVID-19 Model

The IHME model[101] uses the "*curve-fitting/extrapolation*" approach to predict the number of hospitalizations and deaths in the US by state for the next four months. The study uses data on daily confirmed COVID-19 deaths from WHO websites and local and national governments, data on hospital capacity and utilization for US states, and observed COVID-19 utilization data from select locations.

Since the projections are based not on transmission dynamics but on "a statistical model with no epidemiologic basis", the IHME model has some limitations based on its statistical assumptions.

- Assumption of universal social distancing policies: The model assumes systematic variation in mortality curves across regions is captured by timing of social distancing decisions and that other differences are explained by random effects. However, the effects of social distancing policies are not the same everywhere and may not be implemented in all regions.

- Assumption of early relaxation of social distancing: The model assumes stringent social distancing will be in place until deaths drop to below 0.3 per million per capita, and it projects zero deaths in July and August 2020 without virus reintroduction, which proved to be incorrect.

- Large uncertainty bands: The model presents "best guess" on median projections. However, the forecast may have a wide confidence interval when uncertainties arise from inaccurate temporal data on mortality and hospitalization counts, or inaccuracies in assumptions regarding the timing and effect of social distancing policies across regions.

Los Alamos National Laboratory (LANL) COVID-19 Cases and Deaths Forecasting Model

The LANL model[102] uses a "*curve-fitting/extrapolation*" approach to forecast future cases and deaths by U.S. states using assumptions about the growth rate and the presence of social distancing interventions through May 2020.

The model provides state-by-state estimates, so that multiple model predictions could collectively provide more robust situational assessments. However, forecasts for U.S. states with under 100 confirmed cases and/or 10 confirmed deaths are likely to be unstable due to the limited number in the sample.

Additionally, although the model assumes that there will continue to be interventions, it does not specifically assume what those interventions will be. Therefore, the probabilistic approach cannot model effects of specific interventions or be used to create "what-if" scenarios. By assuming various possible interventions, the forecast may result in wider prediction intervals than a model with stricter assumptions. An uncertainty in future trajectories exists given the possibilities of changing intervention strategies, changing case definitions, and changing rates of testing.

Imperial College Non-pharmaceutical Intervention (NPI) Model

The Imperial College NPI model[103] uses the *SEIR* method to project US cases, deaths across a range of different mitigation and suppression scenarios, over the next year. This is a detailed model that goes down to the individual level, estimating contacts with individuals in diverse locations: within a household, at school, at work, and in social settings like shopping.

Although the research team makes assumptions based on real-world observations, some crucial information still remains hidden from the modellers. They don't have a precise measure of disease infectivity even though it is known that each infected individual tends to infect about two others. A reliable testing dataset to see who has been infected without showing symptoms — and so could be moved to the recovered group — would be a major benefit for modellers in general, and might significantly alter the predicted path of the pandemic.

Columbia University COVID-19 Risk Model & Mapping Tool

Columbia's risk model[104] is a county-scale *SEIR* model which provides projections on the number of severe cases, hospitalizations, critical care, ICU use, and deaths under different social distancing scenarios, for 3-week and 6-week periods starting April 2020. To better simulate disease transmission after re-openings, the model leverages public county-to-county work commuting data from the US Census Bureau, and SafeGraph estimates of the reduction of inter-county visitor numbers.

A strength of the model is its ability to produce county-specific projections[105]. This level of geographic granularity helps to simulate disease transmission and medical demand on the local level. The model also displays a dynamic simulation by releasing 4 models with varying assumptions on the reproduction number of an infectious disease: 1) same for the 6-week projection; 3) increase by 5% weekly for the next 2 weeks; 4) decrease by 4% weekly as a seasonal effect. Apart from the forecasts in cases/deaths, the mapping tool also shows data related to populations at high risk of severe COVID-19, including the number of people age 65+, and the numbers of people with underlying health conditions such as diabetes. These efforts facilitate medical planning at a county level.

There are several considerations in model interpretation. First, there is a 2 week delay between infection acquisition and case confirmation. The effects of changes in social distancing and contact patterns over the last 2 weeks on virus transmission have yet to be fully observed by the model. Second, the variety of individual contact behavior, population density, and policy responses highly variable contact behavior, population density, and control measures add great uncertainty to the process of model optimization. Finally, the estimates of medical system capacity are based on healthcare infrastructure data like hospital beds, which does not account for staffing or ventilator supplies. Future analyses could incorporate counts of mechanical ventilators in addition to critical care beds.

COVID-19 Projection Model by Northeastern University, Fogarty International Center, Fred Hutchison Cancer Center, University of Florida, etc.

The study uses Global Epidemic and Mobility Model[106] (GLEAM), an "agent-based", stochastic and spatial epidemic model, to project cases and deaths in the U.S. and by state, under no-mitigation versus "stay-at-home" scenarios. It assumes that each state's current social distancing policies will continue indefinitely.

Using detailed sociodemographic data such as Census data and survey results from publicly available sources, the research team constructed representative synthetic populations for each state in the US. Important population features include age structure, household composition, school structure, and employment rates. The model uses estimates for the effect of school closures, smart working and social distancing effects on the transmissibility of SARS-CoV-2.

As this model heavily relies on the computational simulation of social contact patterns with synthetic population data, it relies on assumptions that cannot take all factors into account. First, the model does not include currently pre-symptomatic transmission. All estimates do not consider the likely introduction of additional mitigation policies issued in states that experience elevated epidemic activity. Second, the modelling estimates for the impact of "stay-at-home" policies use data on contact patterns from Wuhan, China. These results may not be directly generalizable as policy implementation is different in the US. Third, the model only focuses on the impact of social distancing policies instead of other strategies such as contact tracing in reducing transmission. Seasonal drivers that may impact disease transmissibility, such as temperature or humidity, are not considered. Finally, the study does not consider superspreading events and differential transmissibility across age brackets.

REFERENCES

[1] Patil, DJ. "6 Lessons Learned to Get Ready for the Next Wave of COVID." Medium, June 22, 2020. Retrieved from **https://www.belfercenter.org/publication/6-lessons-learned-get-ready-next-wave-covid**.

[2] HHS Protect Public Data Hub, Department of Health and Human Services. Retrieved from **https://protect-public.hhs.gov/pages/hospital-capacity**.

[3] CDC COVID Data Tracker, Centers for Disease Control and Prevention. Retrieved from **https://www.cdc.gov/covid-data-tracker/#cases**.

[4] WHO Coronavirus Disease (COVID-19) Dashboard, World Health Organization. Retrieved from **https://covid19.who.int/**.

[5] COVID-19 Interactive Map, Johns Hopkins University School of Medicine Coronavirus Resource Center (CRC). Retrieved from **https://coronavirus.jhu.edu/us-map**.

[6] COVID-19 Dashboard, Johns Hopkins University Center for Systems Science and Engineering. Retrieved from **https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6**.

[7] Roser, Max, et al. "Coronavirus Pandemic (COVID-19)". OurWorldInData.org. Retrieved from **https://ourworldindata.org/coronavirus**.

[8] "Download today's data on the geographic distribution of COVID-19 cases worldwide." European Centre for Disease Prevention and Control. Retrieved from **https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide**.

[9] The COVID Tracking Project, The Atlantic Monthly Group. Retrieved from **https://covidtracking.com/**.

[10] The COVID Racial Data Tracker, Boston University Center for Antiracist Research. Retrieved from **https://www.bu.edu/antiracism-center/research/datalab/**.

[11] "Coronavirus in the U.S.: Latest Map and Case Count." New York Times. Retrieved from **https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html**.

[12] Killeen, Benjamin D., et al. "A County-level Dataset for Informing the United States' Response to COVID-19." arXiv.org, April 1, 2020. Retrieved from **https://arxiv.org/abs/2004.00756**.

[13] Schneider, Eric C., "Failing the Test — The Tragic Data Gap Undermining the U.S. Pandemic Response." The New England Journal of Medicine, July 23, 2020. Retrieved from **https://www.nejm.org/doi/full/10.1056/NEJMp2014836**.

[14] Kim, Soo Rin, et al. "Want A COVID-19 Test? It's Much Easier To Get In Wealthier, Whiter Neighborhoods." FiveThirtyEight, July 23, 2020. Retrieved from **https://fivethirtyeight.com/features/white-neighborhoods-have-more-access-to-covid-19-testing-sites/**.

[15] Artiga, Samantha, and Hinton, Elizabeth. "Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity." Kaiser Family Foundation, May 10, 2018. Retrieved from **https://www.kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/**.

[16] "Leveraging Data on the Social Determinants of Health - October 2019 Roundtable Report." CODE, December, 2019. Retrieved from **http://reports.opendataenterprise.org/Leveraging-Data-on-SDOH-Summary-Report-FINAL.pdf**.

[17] Artiga, Samantha, and Hinton, Elizabeth. "Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity." Kaiser Family Foundation, May 10, 2018. Retrieved from **https://www.kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/**.

[18] 2019 ACS Data Release New and Notable. The U.S. Census Bureau, June 5, 2020. Retrieved from **https://www.census.gov/programs-surveys/acs/news/data-releases/2019/release.html**.

[19] Gangopadhyaya, Anuj, Garrett, Bowen. "Unemployment, Health Insurance, and the COVID-19 Recession." Robert Wood Johnson Foundation, April 2020. Retrieved from **https://www.urban.org/sites/default/files/publication/101946/unemployment-health-insurance-and-the-covid-19-recession_1.pdf**.

[20] Wiemers, Emily E, et al. "Disparities in Vulnerability to Severe Complications from COVID-19 in the United States." National Institutes of Health, June 22, 2020. Retrieved from **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273290/**.

[21] Unemployment data, U.S. Bureau of Labor Statistics. Retrieved from "The Employment Situation - June 2020." News Release, Bureau of Labor Statistics, July 2, 2020. Retrieved from **https://www.bls.gov/news.release/pdf/empsit.pdf**.

[22] Income Tax Statistics, Internal Revenue Service (IRS). Retrieved from **https://www.irs.gov/statistics**.

[23] Black, Meg, "How COVID-19 Is Impacting People Experiencing Homelessness." Global Citizen, April 18, 2020. Retrieved from **https://www.globalcitizen.org/en/content/coronavirus-impact-on-homessless-population/**.

[24] Walk Score website **https://www.walkscore.com/**.

[25] Housing data, Zillow. Retrieved from **https://www.zillow.com/research/data/**.

[26] Transportation indicators, U.S. Department of Transportation. Retrieved from **https://www.transportation.gov/mission/health/indicators**.

[27] Tools and Data, National Park Service (NPS). Retrieved from **https://www.nps.gov/subjects/gisandmapping/tools-and-data.htm**.

[28] "The Employment Situation." Bureau of Labor Statistics, July 2020. Retrieved from **https://www.bls.gov/news.release/pdf/empsit.pdf**.

[29] Common Core of Data, America's Public Schools. Retrieved from **https://nces.ed.gov/ccd/data_tables.asp#Page:1**.

[30] Selected Social Characteristics in the United States, United States Census Bureau. Retrieved from **https://data.census.gov/cedsci/all?q=dp&g=0400000US01.140000&tid=ACSDP5Y2018.DP02&hidePreview=false&t=Education%3AEducational%20Attainment%3ASchool%20Enrollment&vintage=2018&layer=VT_2018_140_00_PY_D1&cid=DP02_0001E**.

31 "USDA Increases Monthly SNAP Benefits by 40%." U.S. Department of Agriculture, April 22, 2020. Retrieved from **https://www.usda.gov/media/press-releases/2020/04/22/usda-increases-monthly-snap-benefits-40**.

32 Featured Reports, Food and Nutrition Service, U.S. Department of Agriculture. Retrieved from **https://www.fns.usda.gov/pd/overview**.

33 Public Assistance data, American Community Survey, U.S. Census Bureau. Retrieved from **https://data.census.gov/cedsci/table?tid=ACSDT1Y2018.B19058&hidePreview=false&vintage=2018&layer=VT_2018_140_00_PY_D1&cid=DP02_0001E&t=SNAP%2FFood%20Stamps**.

34 De Weerdt, Annick, et al. "Pre-admission air pollution exposure prolongs the duration of ventilation in intensive care patients." Intensive Care Medicine. March 17, 2020. Retrieved from **https://pubmed.ncbi.nlm.nih.gov/32185459/**.

35 Parshley, Lois. "The deadly mix of Covid-19, air pollution, and inequality, explained." Vox. April 11, 2020. Retrieved from **https://www.vox.com/2020/4/11/21217040/coronavirus-in-us-air-pollution-asthma-black-americans**.

36 "Air Data: Air Quality Data Collected at Outdoor Monitors Across the US." United States Environmental Protection Agency. Retrieved from **https://www.epa.gov/outdoor-air-quality-data**.

37 Energy and Power Plants datasets, U.S. Department of Energy (DOE). Retrieved from **https://catalog.data.gov/organization/doe-gov**.

38 Jvion website: **https://jvion.com/**.

39 ZeOmega website: **https://www.zeomega.com/**.

40 Socially Determined website: **https://www.sociallydetermined.com/**.

41 "Hospitalization Surveillance Network COVID-NET." CDC, May 14, 2020. Retrieved from **https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covid-net/purpose-methods.html**.

42 Emerging Infections Program (EIP), CDC. Retrieved from **https://www.cdc.gov/ncezid/dpei/eip/index.html**.

43 The Influenza Hospitalization Surveillance Project (FluSurv - NET), CDC. Retrieved from **https://www.cdc.gov/flu/weekly/influenza-hospitalization-surveillance.htm**.

44 Kim, Lindsay, et al. "Risk Factors for Intensive Care Unit Admission and In-hospital Mortality among Hospitalized Adults Identified through the U.S. Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET)." Clinical Infectious Diseases, July 16, 2020. Retrieved from **https://academic.oup.com/cid/article/doi/10.1093/cid/ciaa1012/5872581**.

45 Garg, Shikha, et al., "Hospitalization Rates and Characteristics of Patient Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020." Morbidity and Mortality Weekly Report (MMWR), April 17, 2020. Retrieved from **https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm**.

46 The COVID-19 Community Vulnerability Index (CCVI) map, Surgo Foundation. Retrieved from **https://precisionforcovid.org/ccvi**.

[47] CDC website: **https://www.cdc.gov/**.

[48] Centers for Medicare & Medicaid Services website: **https://www.cms.gov/**.

[49] Harvard Global Health Institute website: **https://globalhealth.harvard.edu/**.

[50] Policy Map website: **https://www.policymap.com/**.

[51] U.S. Bureau of Labor Statistics website: **https://www.bls.gov/**.

[52] United States Census Bureau website: **https://www.census.gov/**.

[53] Association of Public Health Laboratories website: **https://www.aphl.org/Pages/default.aspx**.

[54] COVID-19 Community Vulnerability Index (CCVI). Retrieved from **https://docs.google.com/spreadsheets/ d/1qEPuziEpxj-VG11IAZoa5RWEr4GhNoxMn7aBdU76O5k/edit#gid=617443512**.

[55] COVID Community Vulnerability Map, Jvion. Retrieved from **https://covid19.jvion.com/**.

[56] Microsoft website: **https://www.microsoft.com/en-us/**.

[57] Johns Hopkins Coronavirus Resource Center website: **https://coronavirus.jhu.edu/map.html**.

[58] COVIDMINDER Rensselaer Polytechnic Institute (RPI) website: **https://covidminder.idea.rpi.edu/?tab=state_ report_cards**.

[59] The Rensselaer Institute for Data Exploration and Applications (IDEA) website: **https://idea.rpi.edu/**.

[60] United Health Foundation website: **https://www.unitedhealthfoundation.org/**.

[61] CDC website: **https://www.cdc.gov/**.

[62] "Health Equity Considerations & Racial & Ethnic Minority Groups." CDC, July 24, 2020. Retrieved from **https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html?CDC_ AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra- precautions%2Fracial-ethnic-minorities.html**.

[63] Godoy, Maria, and Wood, Daniel. "What Do Coronavirus Racial Disparities Look Like State By State?" NPR.org. May 30, 2020. Retrieved from **https://www.npr.org/sections/health-shots/2020/05/30/865413079/ what-do-coronavirus-racial-disparities-look-like-state-by-state**.

[64] "National Infrastructure for Mitigating the Impact of COVID-19 within Racial and Ethnic Minority Communities (NIMIC) Initiative - The Office of Minority Health." Accessed August 7, 2020. **https://www.minorityhealth. hhs.gov/omh/Content.aspx?ID=17500&lvl=2&lvlid=8**.

[65] Goldstein, Amy, "Income emerges as a major predictor of coronavirus infections, along with race." The Washington Post, June 22, 2020. Retrieved from **https://www.washingtonpost.com/health/income- emerges-as-a-major-predictor-of-coronavirus-infections-along-with-race/2020/06/22/9276f31e-b4a3- 11ea-a510-55bf26485c93_story.html**.

[66] "Living in Close Quarters - How to Protect Those That Are Most Vulnerable." CDC, May 29, 2020. Retrieved from **https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/living-in-close-quarters.html**.

67 "Food Desert.", Food Empowerment Project. Retrieved from https://foodispower.org/access-health/food-deserts/.

68 "Profile: American Indian/Alaska Native." U.S. Department of HHS Office of Minority Health. Retrieved from https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=62.

69 Plymyer, David A. "The next battleground: Keeping the elderly in multi-generational households safe from Covid-19." BaltimoreBrew, May 9, 2020. Retrieved from https://baltimorebrew.com/2020/05/09/the-next-battleground-keeping-the-elderly-in-multi-generational-households-safe-from-covid-19/.

70 "Social Distancing - Keep a Safe Distance to Slow the Spread." CDC, July 15, 2020. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html.

71 "Thriving Together: A Springboard Equitable Recovery & Resilience in Communities Across America." Well Being Trust. Retrieved from https://thriving.us/.

72 "Promoting Resilience," March 10, 2020. https://www.usaid.gov/global-health/health-systems-innovation/health-systems/promoting-resilience.

73 "Request for Information-Long-Term Monitoring of Health Care System Resilience." Department of Health and Human Services, June 5, 2020. Retrieved from https://www.federalregister.gov/documents/2020/06/05/2020-12238/request-for-information-long-term-monitoring-of-health-care-system-resilience.

74 Mukherjee, Siddharta. "What the Coronavirus Crisis Reveals About American Medicine | The New Yorker." The New Yorker, April 27, 2020. https://www.newyorker.com/magazine/2020/05/04/what-the-coronavirus-crisis-reveals-about-american-medicine.

75 Fridell, My, et al. "Health System Resilience: What Are We Talking About? A Scoping Review Mapping Characteristics and Keywords." International Journal of Health Policy and Management, September 17, 2019. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6943300/#R1.

76 Manning, Anne, "Planning for a disaster: Model can predict hospital resilience for natural disasters, pandemics." Colorado State University, April 27, 2020. Retrieved from https://engr.source.colostate.edu/planning-for-a-disaster-model-can-predict-hospital-resilience-for-natural-disasters-pandemics/.

77 "Using AI to help health experts address the COVID-19 pandemic." Facebook AI, June 18, 2020. Retrieved from https://ai.facebook.com/blog/using-ai-to-help-health-experts-address-the-covid-19-pandemic.

78 New Jersey COVID-19 Information Hub, Official Site of the State of New Jersey. Retrieved from https://covid19.nj.gov/?fbclid=IwAR0aNYYiZHqmLtS60xzhR6jbsE_R6Q3OPJN1Nb6P0kQPCtuNLzwfgi6KBII.

79 New York State Statewide COVID-19 Testing, New York State Department of Health. Retrieved from https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing/xdss-u53e?fbclid=IwAR0E8kBjZb0DiR1z9bsxesHkDKY-xS3j2-8-6QLmMRNEHfhxfKpWfXBS8Ic.

80 COVID-19 Mobility Data Network, Facebook Data for Good. Retrieved from https://visualization.covid19mobility.org/?date=2020-07-18&dates=2020-04-18_2020-07-18&region=WORLD.

81 Limaye, Vijay, and Knowlton, Kim, "Now's The Time to Strengthen & Modernize Public Health Data." NRDC, July 23, 2020. Retrieved from https://www.nrdc.org/experts/vijay-limaye/nows-time-strengthen-modernize-public-health-data.

[82] Climate-Ready States and Cities Initiative, CDC. Retrieved from **https://www.cdc.gov/climateandhealth/climate_ready.htm#:~:text=CDC's%20Climate%2DReady%20States%20and,impacts%2C%20and%20their%20most%20at%2D**.

[83] **https://teamkyfund.ky.gov/**

[84] "Telehealth Program." Gov Status. Retrieved from **https://govstatus.egov.com/telehealth-program**.

[85] "Louisiana COVID-19 Health Equity Task Force." Southern University System. Retrieved from **http://www.sus.edu/page/louisiana-covid19-health-equity-task-force**.

[86] Sullivan, Jennifer, "Not Answers, Just Leadership." Indiana Family and Social Services Administration, May 19, 2020. Retrieved from **https://www.milbank.org/2020/05/not-answers-just-leadership/**.

[87] Indiana 2-1-1 Community Service. Retrieved from **https://in211.communityos.org/**.

[88] Health Atlas, University of California, San Francisco (UCSF). Retrieved from **https://healthatlas.ucsf.edu/?active&covid=2020-07-23&unemployment=2020-04-01&zoom=6.5&lng=-122.12618778049705&lat=37.64208119839553**.

[89] COVID-19 Interactive Map, John Hopkins University School of Medicine Coronavirus Resource Center (CRC). Retrieved from **https://coronavirus.jhu.edu/us-map**.

[90] COVID-19 Projections, The Institute for Health Metrics and Evaluation (IHME) at University of Washington. Retrieved from **https://covid19.healthdata.org/united-states-of-america**.

[91] LANL COVID-19 Cases and Deaths Forecasts, Los Alamos National Laboratory (LANL). Retrieved from **https://covid-19.bsvgateway.org/**.

[92] Imperial College research team, et al., "Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand." March 16, 2020. Retrieved from **https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/**.

[93] COVID-19 Risk Model & Mapping Tool, Columbia University. Retrieved from **https://columbia.maps.arcgis.com/apps/webappviewer/index.html?id=ade6ba85450c4325a12a5b9c09ba796c**.

[94] COVID-19 Modeling - GLEAM Project, Northeastern University, et al. Retrieved from **https://covid19.gleamproject.org/#about**.

[95] Resolve to Save Lives Initiative, "Essential information for states and counties to publicly report." The Washington Post, July 16, 2020. Retrieved from **https://context-cdn.washingtonpost.com/notes/prod/default/documents/eef52ba6-54a0-4f69-beae-b974322c24d2/note/c6841242-2a06-4950-a4af-dd44e8c0e963**.

[96] Florko, Nicholas, and Boodman, Eric. "How HHS's new hospital data reporting system will actually affect the U.S. Covid-19 response." Statnews, July 16, 2020. Retrieved from **https://www.statnews.com/2020/07/16/hospital-data-reporting-covid-19/**.

[97] Evans, Clare R. "Adding interactions to models of intersectional health inequalities: Comparing multilevel and conventional methods." Social Science & Medicine, January 2019. Retrieved from **https://www.sciencedirect.com/science/article/abs/pii/S0277953618306701?via%3Dihub**.

[98] Don, Aaron van, Cooney, Rebecca E Cooney, and Sabin, Miriam L. "COVID-19 exacerbating inequalities in the US." The Lancet journals, April 18, 2020. Retrieved from **https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30893-X/fulltext**.

[99] COVID Scenario Pipeline Github, Infectious Disease Dynamics Group at Johns Hopkins University. Retrieved from **https://github.com/HopkinsIDD/COVIDScenarioPipeline**.

[100] COVID-19 Daily Updates, California Department of Public Health. Retrieved from **https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Immunization/ncov2019.aspx**.

[101] COVID-19 Projections, The Institute for Health Metrics and Evaluation (IHME) at University of Washington. Retrieved from **https://covid19.healthdata.org/united-states-of-america**.

[102] LANL COVID-19 Cases and Deaths Forecasts, Los Alamos National Laboratory (LANL). Retrieved from **https://covid-19.bsvgateway.org/**.

[103] Imperial College research team, et al., "Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand." March 16, 2020. Retrieved from **https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/**.

[104] COVID-19 Risk Model & Mapping Tool, Columbia University. Retrieved from **https://columbia.maps.arcgis.com/apps/webappviewer/index.html?id=ade6ba85450c4325a12a5b9c09ba796c**.

[105] Columbia University research team, "Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties". April 5, 2020. Retrieved from **https://behcolumbia.files.wordpress.com/2020/04/flattening-the-curve-before-it-flattens-us-20200405b.pdf**.

[106] COVID-19 Modeling - GLEAM Project, Northeastern University, et al. Retrieved from **https://covid19.gleamproject.org/#about**.