Using informatics to guide public health policy during the COVID-19 pandemic in the USA

Jay G. Ronquillo¹, William T. Lester²,³, Diana M. Zuckerman⁴

¹Grinformatics, Haymarket, VA 20169, USA
²Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA 02114, USA
³Harvard Medical School, Boston, MA 02115, USA
⁴National Center for Health Research, Washington, DC 20036, USA

Address correspondence to Jay G. Ronquillo, E-mail: jgr@grinformatics.com

ABSTRACT

Background Current and future pandemics will require informatics solutions to assess the risks, resources and policies to guide better public health decision-making.

Methods Cross-sectional study of all COVID-19 cases and deaths in the USA on a population- and resource-adjusted basis (as of 24 April 2020) by applying biomedical informatics and data visualization tools to several public and federal government datasets, including analysis of the impact of statewide stay-at-home orders.

Results There were 2753.2 cases and 158.0 deaths per million residents, respectively, in the USA with variable distributions throughout divisions, regions and states. Forty-two states and Washington, DC, (84.3%) had statewide stay-at-home orders, with the remaining states having population-adjusted characteristics in the highest risk quartile.

Conclusions Effective national preparedness requires clearly understanding states’ ability to predict, manage and balance public health needs through all stages of a pandemic. This will require leveraging data quickly, correctly and responsibly into sound public health policies.

Keywords Public health

Background

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) presents complex challenges to health professionals, researchers and policymakers.¹ There has been a global effort to make relevant technologies, resources and information available that would accelerate data-driven solutions for all aspects of this pandemic.²,³ However, the current focus on raw counts of cases and deaths in the USA, while necessary, is not sufficient to fully assess the risks, resources and policies to guide better public health decision-making, now and in the future.¹,⁴

Our study analyzes important pandemic characteristics in the USA on a population- and resource-adjusted basis using several publicly available datasets and visualization tools in order to provide deeper insight into critical issues for the current pandemic.

Methods

Study population, data collection and definitions

This was a cross-sectional study of daily report and time series data for the USA from the study date (24 April 2020), that was downloaded from the GitHub COVID-19 data repository hosted by the Center for Systems Science and Engineering at Johns Hopkins University in comma-separated values file format.⁵ Specific fields from the daily report file used for analysis include the name of the state (Province_State), country (Country_Region), total number of COVID-19 cases (Confirmed) and total number of COVID-19 deaths (Deaths). Similarly, specific fields from the time series file include the name of the state (Province_State) and country

Jay G. Ronquillo, MD, MPH, MMSc, MEng
William T. Lester, MD, MS
Diana M. Zuckerman, PhD
(Country Region), along with a single field for each date from 22 January 2020 up to the study date 24 April 2020 for confirmed COVID-19 cases and deaths.

Population-adjusted characteristics were calculated by dividing US state-level totals for COVID-19 (i) cases and (ii) deaths, respectively, by 2019 state population estimates from the US Census Bureau (https://data.census.gov/). Resource-adjusted characteristics were calculated by dividing state-level cases by (i) estimated state-level physician totals from the Agency for Healthcare Research and Quality 2018 Compendium of US Health Systems (https://www.ahrq.gov/chsp), and (ii) published state-level estimates for mechanical ventilators as described in the Society of Critical Care Medicine report on US ICU Resource Availability for COVID-19.

Each state in the USA is responsible for setting their own policies regarding pandemic risk mitigation. Using information from available publication and news sources, we identified states with and without stay-at-home or similar nonpharmaceutical intervention (NPI) orders that were implemented statewide as of the study date. For each state with a statewide stay-at-home order, we calculated the number of days between the effective date of the order and (i) the date of a state's first reported case, (ii) the date of a state's first reported death and (iii) the study date.

Statistical analysis and data visualization
Summary statistics were collected as medians with interquartile range (IQR). Quartiles for COVID-19 case and death characteristics were calculated for each state and visualized as choropleth maps generated in Plotly (Version 4.5.2). All data were integrated using an informatics pipeline built via a Jupyter notebook running Python (Release 3.7.6), and analyses were performed using Microsoft Excel Version 16.16.19 (Redmond, WA).

Results
As of 24 April 2020, there were 903 696 reported COVID-19 cases and 51 859 deaths in the USA. Broken down by US Census region, there were 133 094 cases and 7367 deaths in the midwest, 499 110 cases and 33 510 deaths in the northeast, 182 254 cases and 7196 deaths in the south and 89 238 cases and 3786 deaths in the West. Overall, there were 2753.2 cases and 158.0 deaths per million residents, respectively. Population- and resource-adjusted characteristics by US Census region and division are shown in Table 1. Trends in cases and deaths adjusted for state populations and resource estimates are visualized in Figure 1.

Forty-two states and Washington, DC, (84.3%) had statewide stay-at-home orders for all residents, implemented at a median (IQR) of 22.0 (15.5–27.0) days after first reported case and 8.0 (4.0–14.5) days after first death, and have been in place for a median of 28.0 (23.5–31.0) days as of the study date. At the time that their respective orders became effective, states had a median of 168.2 (88.4–310.4) cases per million residents and 2.5 (0.8–6.8) deaths per million residents. At these thresholds, all states without statewide orders (Arkansas, Iowa, Nebraska, North Dakota, Oklahoma, South Dakota, Utah, Wyoming) would be ranked in the highest quartile for cases per million residents, as well as in the highest quartile for deaths per million residents.

Discussion
Main finding of this study
Most statewide stay-at-home orders were implemented roughly 2–4 weeks after a state's outbreak was first detected, and have only been in place for a few weeks at the time this study was performed. The 1918 influenza pandemic triggered nonpharmaceutical intervention orders in many US cities lasting several months, had multiple (e.g. two or more) waves of pandemic infections, and killed more than half a million people in the USA and tens of millions worldwide. Yet even for that pandemic, US cities with NPI orders that were (i) started soon after an outbreak was detected, (ii) longer in duration, and (iii) broader in scope had better overall outcomes than cities without those characteristics. It remains concerning that most states currently without stay-at-home orders have population-adjusted case and death metrics that are just as grave as the rest of the country. We strongly recommend that all states implement or continue to implement responsible mitigation strategies focused on protecting their most vulnerable populations.

What is already known on this topic
The US Food and Drug Administration (FDA) is responsible for regulating medical devices and laboratory developed tests (LDTs) essential to address the current pandemic, including products like mechanical ventilators and diagnostic tests for COVID-19. The traditional regulatory approval process requires that companies provide evidence of a product's safety, effectiveness and performance; however, the FDA recently issued several Emergency Use Authorizations which lower regulatory standards in order to quickly address urgent supply shortages. As the pandemic continues to intensify, it will be critical to improve the quality of all medical devices (including LDTs) reaching patients. COVID-19 diagnostic
Table 1 Population- and resource-adjusted characteristics of reported COVID-19 cases and deaths by US region and division, as of 24 April 2020

| US region and division | Number of cases per million residents | Number of deaths per million residents | Number of cases per hundred physicians | Number of cases per hundred ventilators |
|------------------------|--------------------------------------|----------------------------------------|---------------------------------------|----------------------------------------|
| Northeast              |                                      |                                        |                                       |                                        |
| New England            | 5733.6                               | 314.5                                  | 215.3                                 | 3036.6                                 |
| Middle Atlantic        | 10063.6                              | 701.1                                  | 449.3                                 | 4596.9                                 |
| Total                  | 8915.4                               | 598.6                                  | 379.1                                 | 4226.5                                 |
| Midwest                |                                      |                                        |                                       |                                        |
| East North Central     | 2356.1                               | 140.2                                  | 124.2                                 | 1198.1                                 |
| West North Central     | 1054.3                               | 37.0                                   | 39.4                                  | 551.1                                  |
| Total                  | 1947.8                               | 107.8                                  | 91.0                                  | 999.1                                  |
| South                  |                                      |                                        |                                       |                                        |
| South Atlantic         | 1559.2                               | 59.1                                   | 113.0                                 | 812.3                                  |
| East South Central     | 1249.8                               | 41.0                                   | 57.5                                  | 576.8                                  |
| West South Central     | 1371.6                               | 62.0                                   | 127.1                                 | 705.1                                  |
| Total                  | 1451.3                               | 57.3                                   | 103.3                                 | 738.4                                  |
| West                   |                                      |                                        |                                       |                                        |
| Mountain               | 1279.0                               | 53.7                                   | 123.6                                 | 739.1                                  |
| Pacific                | 1074.0                               | 45.8                                   | 51.9                                  | 694.4                                  |
| Total                  | 1139.0                               | 48.3                                   | 65.4                                  | 709.7                                  |

Tests with low specificity (and high false positive rates) could lead to unnecessary quarantines, mental stress and wasted hospital resources while low sensitivity (high false negative) tests could lead to multiple waves of uncontrolled community transmission. Until there is a robust supply of accurate, validated diagnostic tests to gauge community spread, the high-risk (upper quartile) states in our figure highlight where greater healthcare resource capacity is needed to prevent overwhelmed hospital systems from increasing patient mortality and putting healthcare employees at risk of infection.1,7,18,19

What this study adds

Effective national preparedness requires clearly understanding states’ ability to predict, manage and balance public health needs through all stages of a pandemic.1,10,20 The rapid spread of SARS-CoV-2 has exposed the limited availability of key resources, from personal protective equipment to mechanical ventilators to the diverse healthcare providers at the front lines of clinical care.1 Looking beyond raw case and death counts by adjusting for publicly accessible data on populations and resource estimates can help clarify risks and inform public health policy.5,21 Choropleth maps, for example, can help policymakers visually gauge which states are at risk for negative outcomes, where public health strategies should shift from containment to risk mitigation, and for how long policies should remain in place.1,10,18

Limitations of this study

Our study had several limitations. First, the numbers of COVID-19 cases and deaths are likely underestimated, given the current shortage of adequate diagnostic tests across the USA. Second, provider and ventilator estimates are not updated as frequently as the pandemic counts, and thus only provide general guidance on resource availability. Third, our study focuses primarily on state-level data but future research should leverage finer levels of granularity, including data about counties, population density, race, age and social determinants of health. Finally, our work integrates and harmonizes pandemic reports with fragmented data from various federal agencies and published outlets. However, effective public health solutions for COVID-19 and future pandemics will require access to interoperable public health data at all levels.22

Conclusion

The COVID-19 pandemic is neither the first nor last major health challenge that the world will face but lasting success...
will depend on synthesizing data and information quickly, correctly and responsibly into sound national and international public health policies.

**Acknowledgements**

None.

**Authors’ contributions**

All authors included in the manuscript provided substantial contribution to (i) conception and design, acquisition of data, or analysis and interpretation of data, (ii) drafting the article or revising it critically for important intellectual content and (iii) final approval of the completed manuscript. JGR had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; or decision to submit the manuscript for publication.

**Funding**

None.
Conflict of interest

Dr. Ronquillo reports working for Syapse, which had no role in the study, has received cloud research grants from Google and Microsoft during previous work as a medical school faculty member and has received cloud research funding from the Google Cloud for Startups Program. The authors have no other competing interests to declare.

References

1 Adalja AA, Toner E, Inglesby TV. Priorities for the US health community responding to COVID-19. JAMA. 2020; doi: 10.1001/jama.2020.3413.
2 Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/SARS-CoV-2 epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr. 2020;19(8). doi: 10.1186/s12942-020-00202-8.
3 Keesara S, Jonas A, Schulman K. Covid-19 and health care’s digital revolution. N Engl J Med. 2020;382(23):e82. doi: 10.1056/NEJMep2005835.
4 Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19—studies needed. N Engl J Med. 2020. doi: 10.1056/NEJMep2002125.
5 Dong F, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020. doi: 10.1016/S1473-3099(20)30120-1.
6 U.S. Census Bureau. Population, population change, and estimated components of population change: April 1, 2010 to July 1, 2019 (NST-EST2019-alldata). 2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html (26 March 2020, date last accessed).
7 Rubinso L, Vaughn F, Nelson S et al. Mechanical ventilators in US acute care hospitals. Disaster Med Public Health Prep. 2010;4(3):199-206.
8 Halpern NA, Tan KS, SCCM ventilator taskforce. U.S. ICU resource availability for COVID-19. 2020. https://www.sccm.org/Blog/March-2020/United-States-Resource-Availability-for-COVID-19 (20 April 2020, date last accessed).
9 Agency for Healthcare Research and Quality. Compendium of U.S. health systems, 2018. https://www.ahrq.gov/chsp/data-resources/compendium-2018.html (26 March 2020, date last accessed).
10 Hatchett RJ, Mecher CE, Lipsitch M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. Proc Natl Acad Sci USA. 2007;104(18):7582-7.
11 Mervosh S, Lu D. See Which States and Cities Have Told Residents to Stay at Home. The New York Times, 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html (20 April 2020, date last accessed).
12 Haftaje RL, Mello MM. Thinking globally, acting locally—the U.S. response to Covid-19. N Engl J Med. 2020;382(22):e75. doi: 10.1056/NEJMep2006740.
13 Markel H, Lipman HB, Navarro JA et al. Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. JAMA. 2007;298(6):644-54.
14 Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. Nature. 2004;432(7019):904-6.
15 Ronquillo JG, Zuckerman DM. Software-related recalls of health and cyber security: national trends in data breaches of protected health information. Milbank Q. 2017;95(3):535-53.
16 U.S. Food and Drug Administration. Policy for diagnostic tests for coronavirus Disease-2019 during the public health emergency: immediately in effect guidance for clinical laboratories, commercial manufacturers, and Food and Drug Administration staff. 2020. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-diagnostic-tests-coronavirus-disease-2019-during-public-health-emergency (18 April 2020, date last accessed).
17 U.S. food and drug administration. Guidance for industry and FDA staff: statistical guidance on reporting results from studies evaluating diagnostic tests. 2007. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda (18 April 2020, date last accessed).
18 Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. N Engl J Med. 2020;382(13):1268-9.
19 Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health. 2020;8(4):e480.
20 Gostin LO, Wiley LF. Governmental public health powers during the COVID-19 pandemic stay-at-home orders, business closures, and travel restrictions. JAMA. 2020. doi: 10.1001/jama.2020.5460.
21 Ronquillo JG, Winterholler JE, Cwikla K et al. Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information. JAMA Open. 2018;1(1):15-9.
22 Dixon B, Kharrazi H, Lehmann H. Public health and epidemiology informatics: recent research and trends in the United States. Yearb Med Inform. 2015;10:199-206.