Can Comorbidity Data Explain Cross-State and Cross-National Difference in COVID-19 Death Rates?

Jeffrey C Cegan1
Benjamin D Trump1
Susan M Cibulsky2
Zachary A Collier3
Christopher L Cummings4
Scott L Greer5
Holly Jarman5
Kasia Klasa1,5
Gary Kleinman2
Melissa A Surette6
Emily Wells1
Igor Linkov1

1US Army Engineer Research and Development Center, US Army Corps of Engineers, Vicksburg, MS, USA; 2US Department of Health and Human Services, Office of the Assistant Secretary for Preparedness and Response, Boston, MA, USA; 3Radford University, Davis College of Business and Economics, Department of Management, Radford, VA, USA; 4North Carolina State University, Genetic Engineering and Society Center, Raleigh, NC, USA; 5University of Michigan, School of Public Health, Department of Health Management and Policy, Ann Arbor, MI, USA; 6Federal Emergency Management Agency, Region I, Boston, MA, USA

Abstract: Many efforts to predict the impact of COVID-19 on hospitalization, intensive care unit (ICU) utilization, and mortality rely on age and comorbidities. These predictions are foundational to learning, policymaking, and planning for the pandemic, and therefore understanding the relationship between age, comorbidities, and health outcomes is critical to assessing and managing public health risks. From a US government database of 1.4 million patient records collected in May 2020, we extracted the relationships between age and number of comorbidities at the individual level to predict the likelihood of hospitalization, admission to intensive care, and death. We then applied the relationships to each US state and a selection of different countries in order to see whether they predicted observed outcome rates. We found that age and comorbidity data within these geographical regions do not explain much of the international or within-country variation in hospitalization, admission to intensive care, and death. Identifying alternative explanations for the limited predictive power of comorbidities and age at the population level should be considered for future research.

Keywords: comorbidity, health outcomes, COVID-19, mortality rates

Introduction

As of January 2021, the novel coronavirus (COVID-19) retains considerable uncertainty in its pathology despite emerging over a year ago. As the disease envelops the globe and news strains develop, it continues to pose a unique challenge for risk analysts, public health professionals, and policy makers. After its initial outbreak in Wuhan, China, medical and public health analysts identified risk factors for more severe presentation of COVID-19 (eg, hospitalization, admission to an intensive care unit, and fatality).1 Among these risk factors are underlying health conditions, age, race/ethnicity, and socioeconomic and behavioral factors.1 Such indicators are both intuitive and evidence-based, and can frame both short-term and long-term needs related to treatment,2 risk-based responses and policy making.3

However, such clinical guidance does not necessarily answer a fundamental question remaining for pandemic response: given an outbreak of a certain time-frame and magnitude, what is the range of population-level outcomes that countries should anticipate given the underlying health conditions of its population? After all, if analysts cannot answer basic risk analysis questions like “what can go wrong”, “how likely is it”, and “what are the consequences”, the efficacy of planning and response activities will be greatly diminished.4,5 An intuitive response to such...
a question is to analyze the prevalence of underlying health conditions within a population, also referred to as comorbidities, against existing COVID-19 epidemiological trends and health outcomes. By tracking noted comorbidities (eg, heart disease, lung disease, diabetes, cancer, etc.), population pyramids, and a growing set of COVID-19 hospitalization and fatality data, it seems likely that various statistical exercises should be able to roughly confirm comorbidities and age as the core determining factor behind the more severe health outcomes of COVID-19.

We test this intuition in two stages. First, we use a large data set of 1.4 million diagnosed COVID-19 cases (described below) to identify relationships between age, number of relevant comorbidities, and the likelihood of hospitalization, ICU admission, placement on a ventilator, and death. If comorbidities and age profile explain COVID-19 outcomes on the population level, then the relationships identified in a database of this size should provide a guide to population-level outcomes. In our second step, we use the relationships found in the data to predict hospitalization, ICU admission, and death for US states and a selection of countries. We find that comorbidities and age structure data for COVID-19 patients do not predict observed population-scale outcomes. We conclude that something more than a simple count of comorbidities and age structure is needed to predict what COVID-19 will do to populations and health systems.

**COVID-19 and Underlying Health Conditions**

Early reports from China, Iran, and Italy note that risk factors for severe incidence and increased fatality of the disease include those aged ≥65 years—particularly among those ≥85 years—and individuals with at least one comorbidity, 6-9 ie, the occurrence of one or more “medical conditions additional to an index disease”10 which may increase medical complications in inpatient management needs.

Many potential comorbidities have been identified by epidemiological studies around the world as exacerbating the symptoms and adverse health outcomes associated with a COVID-19 infection.11 A meta-analysis of seven studies in China noted that severe disease and increased mortality were associated most prominently with hypertension, diabetes and cardiovascular disease.12 Other studies have found increased risk due to smoking and obesity.6,7 The US CDC notes higher risks for severe illness and fatality associated with various factors including: chronic lung disease (inclusive of asthma, chronic obstructive pulmonary disease [COPD], and emphysema); diabetes mellitus; cardiovascular disease; chronic renal disease; chronic liver disease; immunocompromised condition; neurologic disorder, neurodevelopmental, or intellectual disability; pregnancy; current smoking status; former smoking status; or other chronic disease.13,14

While the relationship between comorbidities and COVID-19 health outcomes has been independently explored for several cities and countries, there is currently limited research that comparatively evaluates the relationship between comorbidities and COVID-19 outcomes internationally. Using historical hospital data, Ioannidis et al15 assessed population-level COVID-19 mortality risks for elderly versus non-elderly individuals without comorbidities in France, Italy, Netherlands, Sweden, and the United States (specifically in the State of Georgia and New York City). They found that, as April 24, 2020, the proportion of these COVID-19 fatalities for individuals who were <65 years old without comorbidities ranged from 0.6% to 2.6% of all COVID-19 deaths. Overall, Ioannidis et al found that the risk of death is 14 to 84 times lower in non-elderly people aged <65 years old, and that the age-dependent risk gradient is modestly sharper in European countries and Canada versus most of the US locations.15 Though Ioannidis et al provide insights that could be used to develop strategies that focus on protecting high-risk elderly individuals during the pandemic, their international comparative work purposefully excludes individuals with comorbidities due to there being “… different definitions of eligible comorbidities and data collection methods”.15

Comorbidities tend to be more prevalent in older populations. Thus, when using an international comparative approach, the distribution of comorbidities across each country’s unique population distribution by age should be considered. Based on Italian COVID-19 mortality data, Gjerstad et al found that after controlling for comorbidities, age had a minor effect on COVID-19 mortality.8 This demonstrates how critical it is to consider comorbidities and comorbidity prevalence. Other research has shown higher fatality rates in developed countries that have longer average lifespans, indicating how comorbidities that are more prevalent at older ages may be related to increased COVID-19 severity. Vanella et al found that case
fatality risk differed based on age structures of known COVID-19 cases and on the number of intensive care beds available and occupied in these countries.\textsuperscript{16}

Sornette et al assessed COVID-19 mortality data across over 50 countries, classified into four groups: (i) Western countries, (ii) East Block and developed South East Asian countries, (iii) Northern Hemisphere developing countries, and (iv) Southern Hemisphere countries.\textsuperscript{17} They analyzed the number of deaths per million across the countries, revealing that Western countries exhibited higher mortality rates. They assessed the relationship between deaths per million and life expectancy, finding a positive relationship between life expectancy of a country and its COVID-19 mortality rate. Though Sornette et al did not focus on comorbidities within their formal analysis, they note that heart disease and respiratory disorders were common across fatal COVID-19 cases across countries. For those who were younger than 65 years old, those who were clinically obese were most at risk.\textsuperscript{17}

To date, literature in this area is limited to epidemiological review of COVID-19 patients and outcomes. Such retrospective reviews prove vital for assessing the relative risk ratios of different comorbidities to provide historical explanation and description, but they cannot provide anticipatory insights given distinct health disparities and comorbidity prevalence across regions and countries. Our assessment provides robust comparison and prediction of severe cases and anticipated fatality rates associated with COVID-19. Rather than review documented COVID-19 caseloads, we utilize historical COVID-19 comorbidity fatality rates along with population health comorbidity survey data to build predictive models that demonstrate the anticipated burden of disease across regions and countries.

**Methods**

COVID-19 health outcomes were collected from the HHS Protect Government Data Hub database from the Centers for Disease Control and Prevention on May 27, 2020, and all subsequent analyses were performed using version 4.0.2 of the R programming language (R Project for Statistical Computing; R Foundation). The database reports on confirmed COVID-19 infected individuals, their underlying health conditions, and resulting hospitalization and mortality outcomes of infected individuals. All data was anonymized and did not contain identifying patient information. The data were gathered through surveying 1.4M individual confirmed cases and the characteristics of each infected person across various participating municipalities throughout the United States. There is roughly an even split of genders among the confirmed cases, with 51% of them as female. While the majority of US states are represented in the data, most cases (30%) are from New York, and the second largest representation of cases (11%) are from New Jersey.

To measure how COVID-19 is affecting various age-comorbidity groups in the United States, we categorized confirmed cases according to the individual’s age group and number of comorbidities. Age groups were defined as 0–17, 18–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80+, and the comorbidities considered include diabetes, cardiovascular disease, chronic lung disease, and renal disease. These comorbidities were chosen based on expert opinion about their relevance to COVID-related health outcomes and their availability in both the HHS Palantir and Behavioral Risk Factor Surveillance System (BRFSS) datasets, the latter applied later in the analysis. Cases were categorized as having 0, 1, or 2+ of these comorbidities and by age group. Many of the records were missing critical information, in particular information related to specific health outcomes and the presence of a comorbidity. Due to the nature of how the surveys were conducted, we assumed that blank comorbidity or health outcome responses were negative, meaning the individual did not have that specific comorbidity or health outcome.

The data allowed us to measure the change in health outcomes based on an individual’s age and number of comorbidities. Hospitalization, intensive care, mechanical ventilation, and death rates were calculated by age and comorbidity groupings, as shown in Table 1. The percentages in Table 1 can then be used to help predict the outcomes of COVID-19 among the infected based on population characteristics, ie, quantifying how much worse the hospitalization and mortality rates would be in an older population or one with more comorbidities.

To create age-specific comorbidity prevalence rates for the United States, we used data from the US Census Bureau and the Centers for Disease Control and Prevention (CDC) BRFSS.\textsuperscript{18} First, state population data came from the US Census Bureau’s National Center for Health Statistics bridged-race population estimates of the July 1st resident population, specifically the bridged-race Vintage 2018 (2010–2018) postcensal population estimates.\textsuperscript{19} These estimates were produced by the US
Census Bureau in collaboration with the National Center for Health Statistics (NCHS).

Finally, we predict death rates across states in the United States (using the Palantir HHS COVID-19 health outcomes) as well as select international countries by applying age- and comorbidity-specific death rates from Table 1 to the population profiles in a given state or country, respectively. Our international comparison was enabled by two sources. First, we used established population-level comorbidity prevalence data within 28 countries. Their investigation was based from the World Health Organization World Health Survey Data Archive collected through cross-sectional national survey reports conducted between 2002 and 2004. Therefore, we only included countries and comorbidities analyzed in this study. We then updated population age parameters to current data for each of the 28 countries using the 2020 Central Intelligence Agency World Factbook which prioritizes updating world population data as collected from the Bureau of the Census as well as the United Nations Population Division. Together, these datasets facilitate contemporary international comparisons that provide empirical estimates of chronic disease and multimorbidity prevalence across the lifespan for all 28 countries.

### Results

The results of the analysis are shown in Figure 1: Panel A for the US states, and in Figure 1: Panel B for several nations. Panels A and B show what happens when the likelihoods presented in Table 1 are applied to populations with different levels of comorbidity and demographic profiles, ceteris paribus. In Figure 1: Panel A, it can be seen that the rates of hospitalization, admittance to the ICU, placement on a ventilator, and death for infected people would be highest in West Virginia, which is has the highest rate of cardiovascular disease in the United States, whereas they are the lowest in Utah, one of the youngest and healthiest states in the nation. In Figure 1: Panel B, data for European nations tend to show higher rates of hospitalization, etc., while Asian and African nations show below-average rates.

To understand the effectiveness of this model we then compared predicted death rates with documented death rates, calculated as the number of deaths divided

### Table 1 Age-Comorbidity Group Rates for COVID-19 Health Outcomes According to the Dataset

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Comorbidities</th>
<th>Hospital Rate</th>
<th>ICU Rate</th>
<th>Vent. Rate</th>
<th>Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–17 Years</td>
<td>0</td>
<td>2.37%</td>
<td>0.40%</td>
<td>0.10%</td>
<td>0.04%</td>
</tr>
<tr>
<td>0–17 Years</td>
<td>1</td>
<td>13.24%</td>
<td>3.38%</td>
<td>0.97%</td>
<td>0.39%</td>
</tr>
<tr>
<td>0–17 Years</td>
<td>2+</td>
<td>48.78%</td>
<td>19.51%</td>
<td>13.41%</td>
<td>7.32%</td>
</tr>
<tr>
<td>18–29 Years</td>
<td>0</td>
<td>2.42%</td>
<td>0.36%</td>
<td>0.09%</td>
<td>0.10%</td>
</tr>
<tr>
<td>18–29 Years</td>
<td>1</td>
<td>12.03%</td>
<td>2.22%</td>
<td>1.21%</td>
<td>1.12%</td>
</tr>
<tr>
<td>18–29 Years</td>
<td>2+</td>
<td>48.78%</td>
<td>14.63%</td>
<td>9.15%</td>
<td>7.11%</td>
</tr>
<tr>
<td>30–39 Years</td>
<td>0</td>
<td>4.07%</td>
<td>0.67%</td>
<td>0.19%</td>
<td>0.26%</td>
</tr>
<tr>
<td>30–39 Years</td>
<td>1</td>
<td>20.08%</td>
<td>4.02%</td>
<td>2.34%</td>
<td>2.23%</td>
</tr>
<tr>
<td>30–39 Years</td>
<td>2+</td>
<td>45.14%</td>
<td>13.52%</td>
<td>8.52%</td>
<td>9.65%</td>
</tr>
<tr>
<td>40–49 Years</td>
<td>0</td>
<td>6.07%</td>
<td>1.16%</td>
<td>0.41%</td>
<td>0.61%</td>
</tr>
<tr>
<td>40–49 Years</td>
<td>1</td>
<td>24.65%</td>
<td>5.09%</td>
<td>3.26%</td>
<td>3.51%</td>
</tr>
<tr>
<td>40–49 Years</td>
<td>2+</td>
<td>49.57%</td>
<td>13.88%</td>
<td>9.60%</td>
<td>11.80%</td>
</tr>
<tr>
<td>50–59 Years</td>
<td>0</td>
<td>8.54%</td>
<td>1.77%</td>
<td>0.67%</td>
<td>1.47%</td>
</tr>
<tr>
<td>50–59 Years</td>
<td>1</td>
<td>28.86%</td>
<td>6.37%</td>
<td>4.37%</td>
<td>5.78%</td>
</tr>
<tr>
<td>50–59 Years</td>
<td>2+</td>
<td>55.69%</td>
<td>16.33%</td>
<td>12.00%</td>
<td>16.47%</td>
</tr>
<tr>
<td>60–69 Years</td>
<td>0</td>
<td>13.45%</td>
<td>2.75%</td>
<td>1.14%</td>
<td>4.07%</td>
</tr>
<tr>
<td>60–69 Years</td>
<td>1</td>
<td>39.62%</td>
<td>8.99%</td>
<td>6.40%</td>
<td>12.71%</td>
</tr>
<tr>
<td>60–69 Years</td>
<td>2+</td>
<td>66.92%</td>
<td>19.48%</td>
<td>14.52%</td>
<td>26.26%</td>
</tr>
<tr>
<td>70–79 Years</td>
<td>0</td>
<td>21.28%</td>
<td>3.94%</td>
<td>1.60%</td>
<td>11.14%</td>
</tr>
<tr>
<td>70–79 Years</td>
<td>1</td>
<td>55.29%</td>
<td>11.32%</td>
<td>7.88%</td>
<td>27.76%</td>
</tr>
<tr>
<td>70–79 Years</td>
<td>2+</td>
<td>74.65%</td>
<td>19.68%</td>
<td>14.45%</td>
<td>39.19%</td>
</tr>
<tr>
<td>80+ Years</td>
<td>0</td>
<td>20.88%</td>
<td>2.61%</td>
<td>0.85%</td>
<td>21.12%</td>
</tr>
<tr>
<td>80+ Years</td>
<td>1</td>
<td>55.48%</td>
<td>7.13%</td>
<td>4.54%</td>
<td>48.84%</td>
</tr>
<tr>
<td>80+ Years</td>
<td>2+</td>
<td>69.79%</td>
<td>12.14%</td>
<td>7.89%</td>
<td>52.74%</td>
</tr>
</tbody>
</table>

Source: HHS Protect Government Data Hub; Authors’ own.
by the number of cases based on data \(^{24}\) as of June 12, 2020 (see Figure 2: Panel A and Panel B). For the US states, we can see that the predicted death rates based on age and comorbidities are higher than actual death rates, sometimes by several percentage points. The Pearson correlation coefficient for US states is \(-0.015\). For the international comparison, there is also a slight systematic over-prediction, with a Pearson correlation coefficient of 0.515.

### Discussion

While our initial results imply that the age distribution of a population and the prevalence of comorbidities can explain some of the differences in health outcomes, additional confounding factors remain. During initial pandemic response, many governments chose to solely focus on age and comorbidities to predict spread of the virus. But, we argue that a simple count of comorbidities and age structure appears to be insufficient to capture the full scope of
population dynamics needed to predict outcomes for communities and health systems. For future pandemics or other health disasters, variables outside of age and comorbidities (eg, social determinants of health, health system capacity, inequities and vulnerable communities) should thus be seriously considered.

The distribution of comorbidities varies widely across regions and between countries, potentially reflecting socio-political factors, health equity issues, and environmental threats. These factors can skew the effect of specific comorbidities on death rates. Future risk analysis studies will need to consider such upstream social determinants of health, as well as the unique health care systems, policies, and institutions in place in specific countries. Banik et al assessed factors that determined COVID-19 fatality rates across 29 developing and developed economies, while considering public health infrastructure, vaccination policies, demographics, and policy intervention. They found that the public health system in combination with population age distribution were powerful factors in determining fatality rates. Additionally, other unpredictable factors may influence disease outcome: researchers are exploring whether the number of infectious viral particles to which people are exposed at the point of infection (the “inoculum”) affects COVID-19 infection severity, as has been shown in Syrian hamsters and ferrets. Other independent variables could include population density, family size, and employment. Additionally, there are social determinants that have been demonstrated to raise risk tolerance in hurricane contexts: individualistic behavior, and a lack of trust in authorities (government, scientists) including to the point of believing oneself capable of gathering sufficient information to know better than official evacuation recommendations.

In aggregate, the number of independent variables in predicting disease outcome may prevent reliable relationships from emerging from large-scale data specific to comorbidities. While comorbidities and age are two important variables that can answer “what can go wrong” during a pandemic, they do not necessarily answer “how likely is it” or “what are the consequences” without the addition of other variables such as socio-economic factors, policies, healthcare capacity, and other determinants of health. Future risk analyses used in predicting pandemic response will require taking a complex systems perspective to increase the efficacy of planning and response activities.

**Conclusion**

We find that data relating to comorbidities and age structure for COVID-19 patients do not predict observed population-scale outcomes. The variation in prevalence of multiple comorbidities within distinct populations may be a prominent factor as to why some regions and countries have distinct health outcomes from others during the pandemic. Studies have demonstrated increased risks exist for patients with substance abuse disorders, HIV and tuberculosis infections, mental health disorders, and specific genetics, among other comorbidities, but not how they interact with each other or other factors. Nor did we find...
studies examining whether individuals with comorbidities display more compensatory protective behaviors. The combined effects of individuals’ comorbidities, behaviors, and resources cannot be captured with data related to comorbidities alone.

Finally, our study has limitations. Several data constraints may have limited the analysis. In the first step, there is a risk of sample bias in the HHS Protect dataset, as well as potential bias from our decision to count missing data as negative. The structure of questions could have encouraged respondents (in high-pressure contexts) to only answer if the answer was affirmative. In the second step, there are possible limitations in available data (eg, measurement or instrumentation error regarding the consistency and thoroughness of COVID-19 testing and documentation). What we find to be a lack of relationship between the rates in Table 1 and the country or state outcomes could in theory be due to data problems on either side. Last, we were unable to calculate comorbidity rates by age cross-nationally using current data. Such data were difficult to access and were not easily comparable. Therefore, we used existing, comparable, but old data from 2004, potentially leading to an undercount of comorbidities in our international comparison.

Our study has strengths. In June 2020, we were the first to analyze a nascent, national health surveillance dataset. We have found that while intuitively important, and clinically very relevant, analysis of comorbidities in the population using one of the largest datasets available early in the pandemic does not lead to population-level predictions that are useful for planners. While comorbidities can be helpful in estimating risk, they do not constitute a complete picture and should be considered in tandem with other risk factors.

We suggest that future research should warrant an exploration of issues beyond the incidence of comorbidities and age profiles which will better prepare cities, states, and countries to understand their true exposure to COVID-19 health outcomes as additional waves of infection spread. Ongoing research in the field of risk analysis can generate evidence in support of adaptive and science-based policy, and provide guidance for clear and effective risk communication to the public.

Acknowledgments

The authors are grateful to Captain W. Russell Webster, USCG (Ret.), CEM for his leadership and guidance on this project. Review and comments from Drs. David E. Golan and Laura C. Green are greatly appreciated.

Disclosure

Dr Zachary A Collier reports personal fees from US Army ERDC, during the conduct of the study. Professor Holly Jarman reports grants from US Army Engineer Research Development Center, during the conduct of the study. The authors reported no other potential conflicts of interest for this work. The views and opinions expressed in this article are those of the individual authors and not those of the U.S. Army, FEMA, HHS or other sponsor organizations.

References


