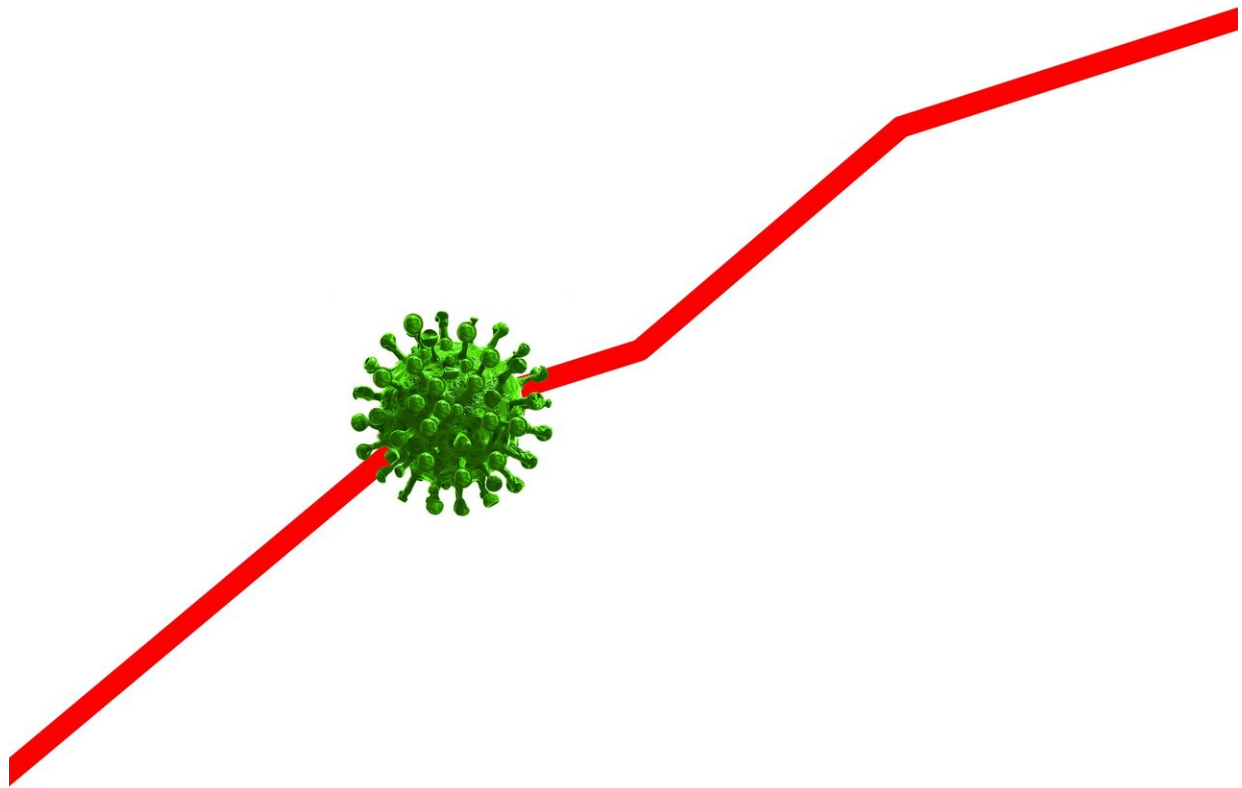# How to use statistics to prepare for the next pandemic

May 18 2021, by R. Alexander Bentley

Credit: CC0 Public Domain

Publicly available statistics about population demographics and culture can help governments prepare for the next pandemic. We have found that by using existing socio-demographic data from early COVID-19 hot

spots, where there was a lot of information, officials could have predicted how COVID-19 would spread through society. The next time there is a global health crisis governments can use our techniques to figure out how a disease will likely move beyond hot spots to regions that are not yet affected.

With a [computational social scientist](link) and a [librarian for science, technology and mathematics research](link), we study the socio-cultural drivers of public health crises, such as obesity. In two peer-reviewed papers that we published in early 2021, which build on our previous research, we analyzed these drivers [at the scale of U.S. counties](link) and [at scale of nations](link). Both studies connected socio-cultural variables to the impact of COVID-19.

For our U.S. study, we collected data from 3,088 U.S. counties on 31 factors that could affect the spread of COVID-19. These factors included population density and ethnicity, commuting habits for work, [voting patterns](link), social connectivity, underlying health conditions and economic information. We collected this information from the U.S. Census Bureau and a variety of other sources.

Using these factors, we built a predictive model of COVID-19 prevalence. We found that just five risk factors can predict between 47% and 60% of variation in COVID-19 prevalence in U.S. counties: population size, population density, public transport, voting patterns and percent African American population. We validated our model by showing that counties which reported fewer COVID-19 cases in April than expected in our model tended to have more cases in July. The results thus provide a new way of discerning when a U.S. county is under-reporting the actual number of infections present in the community.

In the second paper we sought to explain why certain countries, like the U.S., have death tolls in the hundreds of thousands, while other nations

had very few deaths. Using international data from a [large survey](#), measuring [cultural values](#) in 88 countries, we found demographic factors like [population size](#) and obesity levels were important. But more surprising, we found culture was also important, in that open and tolerant societies, as well as those with low trust in institutions, tended to fare the worst.

This analysis made some surprising predictions about the spread of COVID-19 around the world. For example, while many believed in early 2020 that African countries would be heavily affected by COVID-19, our model predicted that they would not. So far this has been true.

In the U.S., which scored high on many of the socio-cultural risk factors—including low trust in institutions, high tolerance toward minorities and high levels of obesity—COVID-19 has hit very hard. Nearly [583,000 people in the U.S. had died from COVID-19](#) as of May 12, 2021. That is the highest absolute number of deaths in any nation so far, and roughly 17.5% of global deaths from the virus, in a country where only 4% of the world population lives.

Governments struggle to predict and plan for the location and extent of disease outbreaks. With so many moving parts, from local mandates like economic shutdowns and face mask recommendations, to international travel bans or restrictions, it seems almost impossible to project the number of cases in different counties or regions. In the average week, how many cases might you expect to have? Should the U.S. expect more cases than Ghana? Why might one city or region be hit harder than another?

We show that additional planning based upon cultural and demographic factors can help predict how outbreaks could progress. It can also reveal which people may be most vulnerable. Properly applied, this data-driven approach might save hundreds of thousands of lives when the next

pandemic hits.

Our goal is to use the predictive power of cultural and [demographic data](link) to anticipate the spread of future pandemics. But neither of our studies specify a relationship between cause and effect.

For example, when looking at the U.S., one of the five predictive factors is the proportion of the [population](link) that is African American: Higher proportions predicted higher infection and death rates. Our analysis, however, did not determine whether this one factor might subsume many other truly causal factors. The social science and public health literature posits reasons why African American populations have suffered more from COVID-19, including bigger households, underlying health conditions and a tendency to work in sectors with greater risk of exposure.

This article is republished from [The Conversation](link) under a Creative Commons license. Read the [original article](link).

Provided by The Conversation