

Credible assumptions replace missing data in COVID analysis

August 6 2020



Credit: Unsplash/CC0 Public Domain

How contagious is COVID-19, and how severe is the virus for those who've caught it?

Everyone wants firm numbers as schools make decisions about in-person versus remote learning, as local and state governments grapple with reopening, and as families care for sick loved ones.

But firm data is missing, said Francesca Molinari, the H.T. Warshow and Robert Irving Warshow Professor in the Department of Economics, in the College of Arts and Sciences. The best way to find out the share of the population that has been exposed to the virus is to either test everyone or to test a random sample of people. But currently not everyone gets tested, and testing is not random; moreover, tests are not perfect. These data challenges have led to wildly divergent predictions in recent months about how many people get infected and how many infected people die.

In research published in the *Journal of Econometrics*, Molinari and Charles F. Manski, the Board of Trustees Professor at Northwestern University, wrote that actual cumulative rates of COVID-19 infection are higher than reported rates of infection, and therefore actual infection fatality rates are lower than reported rates. The researchers reached these conclusions using a technique called "partial identification," which Molinari uses often in her econometrics research.

"You are interested in some quantity, but you cannot learn it exactly," she said. "In this particular instance, we are interested in the infection rate, and we recognize that because we don't have a random sample, we can't learn the exact infection rate from the data."

She and Manski made weak but logical assumptions about COVID-19 data from Illinois, New York and Italy from March 16 to April 24, thereby putting some limits around the incomplete data.

They assumed that the infection rate among those who are tested is higher than the rate among those who are not—a logical assumption

because people showing symptoms are most likely to be tested. The researchers also allowed for the possibility that many negative test results were false—i.e., that the person tested was actually positive but not counted.

These two assumptions drive the actual cumulative infection rates up and push the actual fatality rates down, Molinari said. Cumulative infection rates in New York state as of April 24, according to the researchers, were between 1.7% and 61% of the state's 19.45 million residents (or between 330,650 and 12,020,100 people), with an upper [infection](#) fatality rate of 4.9%. That is substantially lower than the death rate among confirmed infected individuals, which on April 24 was 5.9%.

Infection rates for the same date in Illinois were between 0.04% and 52%; in Italy, they were between 0.06% and 47%.

"The bounds you get are wide," Molinari said, "but they are substantially tighter compared to the bounds you obtain if you assume nothing about the missing data."

Making key assumptions and narrowing the bounds helps policymakers and leaders better understand fatality rates as they try to limit spread of the virus and plan reactivations. Molinari hopes this research will contribute to serious analysis of policies.

Molinari and Manski are working on a follow-up analysis of a longer time period that adds data from California, Florida and Texas to the study.

Provided by Cornell University

Citation: Credible assumptions replace missing data in COVID analysis (2020, August 6)

retrieved 25 April 2024 from

<https://medicalxpress.com/news/2020-08-credible-assumptions-covid-analysis.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.