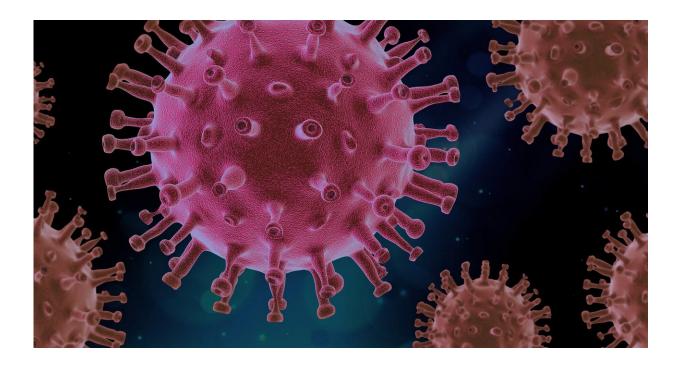# Economist helps solve COVID-19 missing data problems

February 11 2022, by Kate Blackwood

Credit: Pixabay/CC0 Public Domain

When the novel virus SARS-CoV-2 emerged in the early months of 2020, questions about its transmissibility and severity were urgent to public health—but data from test results was limited by questionable accuracy and the non-random nature of test results.

In new research, Jörg Stoye, professor of economics in the College of

Arts and Sciences, found a way to make even limited data sets useful in answering urgent public health questions. Approaching this epidemiological situation as a "[missing data](#) problem" in economics, Stoye determined that limited data provides a range of possible outcomes, and therefore valuable insight to questions of public health.

The article "Bounding infection prevalence by bounding selectivity and accuracy of test: with application to early COVID-19," published in January in *The Econometrics Journal*.

"I investigate what conclusions about the lethality of SARS-CoV-2 could be supported from the scant data available in spring of 2020, and using very weak assumptions," Stoye said. "Evaluated on data from the pandemic's early stage, even the weakest of the novel bounds are reasonably informative. The motivating application is to the COVID-19 pandemic, but the strategy may also be useful elsewhere."

As an economist, Stoye studies partial identification: causality with limited information. "I think a lot about how we can conclude that something causes something else—say, that schooling causes good outcomes in children—even though we don't have an experiment," he said.

Stoye recognized the same problem structure in questions raised by COVID-19.

In spring 2020, he applied partial identification reasoning to the question, hotly debated at the time, of whether COVID-19 was more prevalent or lethal than influenza—and therefore deserving of stronger public health measures.

Two specific limits lay between test result data and truly determining whether COVID-19 was more deadly than the flu. First, test accuracy

was in question. Second, those tested did not represent a random sample of the population; rather, they self-select as people willing to be tested.

Assuming that 2% of the population got tested, Stoye said, and for the sake of argument, all test positive. The range of possible infection rates among the general population could be 2% to 100%, an unhelpful conclusion.

So he introduced a range of boundaries, allowing up to 30% false negatives on PCR tests and also assuming that the untested population could have up to the same incidence as the tested one.

"By April of 2020, these assumptions, while far from pinpointing the disease's lethality, place it well above the lethality of influenza," Stoye said. "To be sure, this was also mainstream scientific opinion at the time, but it was contested by a vocal minority."

This strategy may be useful beyond the COVID-19 pandemic, Stoye said, by determining infection rates of endemic diseases such as influenza. The approach also helps determine efficacy of public health guidelines, such as masking.

The people who mask are not a random selection, he said, and the effect of masking has to do with whether people around you are masked.

Because they work with inherently limited data, economists are on the front lines of researching public health questions, Stoye said.

"Many economists and epidemiologists work on problems that are very similar 'under the hood,'" he said, "Some of the best papers on COVID-19 are collaborations between epidemiology and economics."

 More information: Jörg Stoye, Bounding infection prevalence by

bounding selectivity and accuracy of tests: with application to early COVID-19, *The Econometrics Journal* (2021).