

In Pursuit of Real Coronavirus Numbers

Jillian Mock

April 19, 2020

Editor's note: Find the latest COVID-19 news and guidance in Medscape's [Coronavirus Resource Center](#).

As the COVID-19 pandemic stretches on, there's one big question everyone wants to answer: When it will be safe to get back to some version of normal life? Opening the country back up safely to limit the economic devastation of the shutdown depends on [aggressive testing and tracing](#), which many say the US public health system is not equipped to do. But part of the puzzle is figuring out how many people are — or have been — infected with the virus.

On Friday, [researchers posted a preprint](#) declaring that the number of people infected with SARS-CoV-2 in one California county is far, far higher than previously thought, making the presumed fatality rate much lower. Another recent preprint suggests millions of people across the United States had already contracted COVID-19 in March.

These findings come with appealing implications. In the words of [one article in the *Economist*](#): "If millions of people were infected weeks ago without dying, the virus must be less deadly than official data suggest." Data from these studies supports the [current push](#) to end lockdowns that have strained the world's social and economic fabric, as one [Wall Street Journal opinion columnist wrote](#).

But multiple scientists who took a close look at the California study and shared their thoughts on Twitter said such hopeful findings are misleading, as the analysis does not stand up to close scientific scrutiny. And the researchers behind the second study have already revised their numbers downward to account for data discrepancies.

"We need these kinds of studies and data badly. Unfortunately, this paper is badly misleading (bordering on purposeful?)" wrote [A. Marm Kilpatrick](#), PhD, a zoologist who studies infectious diseases at University of California, Santa Cruz, [on Twitter](#) of the preprint testing the seroprevalence of COVID-19 antibodies in Santa Clara County, California.

"People want to be able to say that the disease is less severe than it is," [Natalie Dean](#), PhD, a biostatistician who studies infectious disease surveillance, surveys, and vaccines at the University of Florida in Gainesville, told *Medscape Medical News*. But Dean, like at least five [other scientists](#) who [took to Twitter](#) to conduct their own peer review of the California study, is very skeptical of the attention-grabbing headlines.

In Santa Clara County, researchers from Stanford University conducted what they say is [the first large-scale community-based COVID-19 prevalence study](#) in a large US county. The scientists recruited patients via Facebook ads targeted by geography and demographics. Participants came to one of three drive-through testing stations, where researchers took a small blood sample and tested it for SARS-CoV-2 antibodies using Premier Biotech's serology test. After testing, the researchers adjusted their findings to account for underrepresented zip codes, sexes, and races/ethnicities, and then adjusted those results again to account for the limitations of the test.

Ultimately, out of 3330 people tested, 50 came back positive for COVID-19 via either IgG or IgM antibodies in the sample. Unadjusted, the seroprevalence in Santa Clara County was 1.5% (exact binomial 95% confidence interval [CI], 1.11 - 1.97%). After weighting the data for demographics and the test characteristics, the researchers determined the population prevalence ranged from 2.49% to 4.16% (95% CI, 1.80 - 3.17% and 95% CI, 2.58 - 5.70%, respectively). This suggests between 48,000 and 81,000 people were infected in Santa Clara County in early April, a 50- to 85-fold increase over the confirmed 956 cases. As the authors write in the study, that corresponds to an infection fatality rate of 0.12% to 0.2%, well below [most estimated fatality rates](#), which range from 4.3% in the US to 13% in Italy.

Among scientists [dissecting the study on Twitter](#), the main criticisms of the paper revolved around a failure to fully account for the imperfections with the antibody test and bias in the population sample, says Dean.

"The only thing they did well was to try to answer a question, but they miserably failed in every aspect of it," says cardiologist [Eric Topol](#), MD, founder and director of the Scripps Research Translational Institute in La Jolla, California, and editor-in-chief of Medscape. (On Twitter, Topol compared the results to [those in Seattle](#), saying those were much lower.)

When a disease is rare in a population, even a really accurate test with a high specificity will turn up a lot of false positives, says Dean. And an accurate antibody assay for COVID-19 is proving extremely challenging to produce, says Topol. Companies and organizations around the world are currently [struggling to deliver speedy and dependable COVID-19 antibody tests](#).

The assay used in the Santa Clara study has not been validated in an extensive number of people, says Topol, suggesting it could be far less accurate than the researchers presumed. The Food and Drug Administration (FDA) has waived its usual

approval process for COVID-19 antibody tests, so assays like this one can be deployed without waiting on FDA validation.

Before testing patients, the researchers ran 67 samples through the test to estimate its sensitivity and specificity; the manufacturer assessed the test using 531 total samples. Pooling the results, the researchers concluded that this particular antibody test had a combined sensitivity of 80.3% (95% CI, 71.1 - 87.0%) and a specificity of 99.5% (95% CI, 98.3 - 99.9%). But as [John Cherian](#), a statistician and biomolecular simulation expert at DE Shaw Research, [pointed out on Twitter](#), if the true specificity of the test was in fact closer to 98.3%, this would mean that almost all of the positive results in the unadjusted 1.5% prevalence rate could potentially be dismissed as false negatives.

Attempts to adjust bias in the sampled population also could have skewed the data, Dean explains. White women, ages 19 to 64, were overrepresented, while Hispanic and Asian populations were underrepresented. As the authors, who were unavailable for comment before press time, acknowledge in the paper, this could be because of their Facebook recruitment method. Not everyone can take off work or even has a car to go to a drive-through testing facility, says Dean.

When the number of positive cases is so small (just 50 total), influential observations can easily muddle the results. For example, if only two people showed up from a particular zip code, and one of them tested positive, that area appears to have an estimated seroprevalence of 50%. "What they did I think was in good faith and it's a reasonable approach, it can just have undesirable characteristics in small surveys," says Dean.

In areas that have high hospitalizations and death, seroprevalence is at about the 10% mark, says Dean. Given that Santa Clara has not been particularly hard-hit by the virus so far, Dean is [skeptical of the 4% seroprevalence finding](#), and would guess the actual seroprevalence could be around 1% or 2%. This would, in turn, imply a higher infection fatality rate than the one cited in the paper.

Tens of Millions?

"In the last weeks, we've seen a flood of studies that are really supporting this view that there are lots of people who have this and we're vastly undercounting," says [Justin Silverman](#), MD, PhD, assistant professor of information science and technology at Penn State University in State College, Pennsylvania.

Silverman and his colleagues recently published their [own preprint that used influenza surveillance data to estimate at least 8.7 million people](#) in the United States were infected with COVID-19 during a three week period in March. In the original manuscript, the researchers published that an estimated 28 million people in the US had COVID-19 during this time period, but [Silverman says](#) they had to revise the number after taking a closer look at some of the underlying data collected by the Centers for Disease Control and Prevention (CDC).

"What we found starting on March 8 on was there was a strong correlation between excess influenza-like illnesses and the path of coronavirus spread across the US," says Silverman.

To produce these results, the researchers extracted data from the CDC's [ILINet database](#), which tracks the number of people presenting with influenza-like symptoms to more than 2600 enrolled providers across the country each week. From this data, Silverman excluded the number of patients with positive [flu](#) tests and then subtracted expected seasonal variation in influenza-like illness (ILI) prevalence, estimated using a model trained on 10 years of ILINet data.

For a 3-week period in March, Silverman and his colleagues assumed nearly all of the remaining, unexplained ILI cases were caused by COVID-19. Finally, the researchers scaled up their results, using the estimated total number of providers per 100,000 residents in each state to estimate the number of unexplained ILI cases nationwide. After publication, the researchers revised the method they used to scale up the findings when they realized the CDC data sometimes counted groups of clinicians as just one provider in the database.

The researchers also estimated the number of COVID-19 cases doubled every 3.5 days, which Silverman says seems to match with the death rate doubling every 3 days so far.

Silverman notes all these statistical findings need to be verified by expanded seroprevalence testing on the ground via studies like the one conducted in Santa Clara.

This study hasn't been discussed as widely on Twitter as the Santa Clara manuscript, but Topol is skeptical. While the US has vastly undercounted the actual number of COVID-19 cases, numbers this high don't make any sense, says Topol. Right now, the total number of confirmed cases worldwide is around 2.3 million, and in the United States confirmed cases are currently over 700,000, according to [tracking data by Johns Hopkins University](#). Unfortunately, we likely will never know how many people were truly infected with COVID-19 because of the slow and flawed rollout of testing, Topol says.

While neither study went through peer review, Topol doesn't think the preprint process, which some are [concerned](#) will allow results to reach the public before they're ready for prime time, is necessarily the problem. After all, the same thing can happen with papers that are peer reviewed, and preprints have allowed the rapid dissemination of new knowledge about COVID-19, which we desperately need right now, he says.

Still, Topol cautions that studies like these can give [the wrong impression](#) about the seriousness of the pandemic and fuel arguments and [protests against social distancing](#), the main tool for slowing the spread of SARS-CoV-2 right now.

And survey studies like the one in Santa Clara can serve an important purpose, says Dean. "I think these quick-and-dirty surveys are very valuable because they provide rapid information, but they have important limitations," she says. "You can't overinterpret a single number."

For more news, follow Medscape on [Facebook](#), [Twitter](#), [Instagram](#), and [YouTube](#).

Medscape Medical News © 2020

Cite this: In Pursuit of Real Coronavirus Numbers - *Medscape* - Apr 19, 2020.