Technology & Ideas

10 Reasons to Doubt the Covid-19 Data

The pandemic's true toll might never be known.

By Cathy O'Neil
14 de abril de 2020 0:15 CEST

Getting counted? *Photographer: John Moore/Getty Images*

Cathy O'Neil is a Bloomberg Opinion columnist. She is a mathematician who has worked as a professor, hedge-fund analyst and data scientist. She founded ORCAA, an algorithmic auditing company, and is the author of "Weapons of Math Destruction."

Read more opinion

Follow @mathbabedotorg on Twitter

If you're like me, you've been watching the daily data on the coronavirus pandemic, seeking glimmers of hope in the trajectories: the infected, the hospitalized, the intubated, the dead.

If only there were more understanding to be had. The more I look at the numbers, the more I see their flaws. Here are my top 10.

1. The number of infected is close to meaningless. Only people who get tested can be counted, and there still aren't enough tests – not even close, and not in any country save perhaps Iceland. The best we can do is estimate how many people are sick by guessing what percentage of the infected can obtain a test. In the U.S., for example, anecdotal evidence suggests that people need to be ill enough to be hospitalized. About 10% of cases merit hospitalization, so the actual number of infected might be about ten times larger than what's reported.

2. The tests aren't accurate and the inaccuracies aren't symmetric. In particular, they produce many more false negatives than false positives – meaning they tend to indicate that people are OK when they're actually sick. Some research suggests that the false negative rate could exceed 30%. This means that estimates of the true number of infections should be once again inflated.

3. The number of tests doesn't equal the number of people tested. Because the tests are so inaccurate, some people get tested twice to be more sure of the results. This means that the share of the population tested compared to the number of people found to be infected paints a rosier picture than reality, offering yet another reason to believe that the actual number of infected is higher.

4. The numbers aren't in sync. People sometimes die weeks after being hospitalized, and they get hospitalized a week or more after testing positive for the virus. So we shouldn't expect the "number of deaths" curve to flatten until pretty long after the "number of cases" curve does. The bright side of this lag is that, since it takes longer to recover than to die, the death rate will go down over time.

5. The meaning of hospitalization is changing. Officials have recently presented flattening hospital admissions as a positive sign. But it takes a lot more to get somebody to the hospital these days. Hotlines are jammed, ambulances are scarce, standards for who gets hospitalized have drastically changed, and people are avoiding overwhelmed emergency rooms. So fewer hospitalizations doesn't necessarily mean that the situation is getting better.

6. Deaths aren't reported immediately or consistently. Various operational issues, such as paper filing and notifying next of kin, determine when a death actually gets registered. This might help explain why the most deaths tend to get reported on Tuesdays. So don't get too excited about good news on a weekend – you might be disappointed by the beginning of the week.

7. Deaths outside hospitals aren't being reported. When people die at home or in nursing facilities, veteran homes, or prisons, they're not always counted. This is a biggie: When France started reporting fatalities in nursing homes, their death count increased by 40%. Belgium reports nursing home deaths pretty well, and they're finding 40% of deaths occur there.

8. The policy for attributing deaths isn't consistent. Once somebody is gone, why waste a valuable test? So doctors might not mention Covid-19 as a contributing cause. It's a judgment call, especially when someone was sick already. This might have a very large effect on the data in certain environments like rehab facilities and nursing homes.

9. Officials may have incentives to hide coronavirus cases. China, Indonesia and Iran have all come under scrutiny for their statistics. "Juking the stats" is not unknown in other contexts in the U.S., either. So don't assume that officials are above outright manipulation.

10. What happens in one place, or on average, might not be applicable everywhere. Some small studies suggest that the Covid-19 mortality rate is about 1% of the infected population. But that doesn't mean it will be the same in the U.S., or in New York City. Specific areas could see much worse death rates, simply because their health care systems are not as comprehensive or their populations have more chronic diseases. The U.S. has plenty of polluted areas that seem to make people more vulnerable to infection and sicker once they get sick. As we've seen in recent days, such disparities are disproportionately affecting people of color.

Appealing as it may be to keep count, the true numbers might not be knowable until much later. Testing needs to be done systematically, even on asymptomatic people. For deaths, precise numbers might never emerge. It's possible to estimate using the number of unexpected deaths compared to a year earlier. But even that's not ideal, because lockdowns might suppress other kinds of deaths – traffic accidents, for example – by forcing people to stay at home.

Don't get me wrong: Watching the official data is not a complete waste of time and attention. The numbers can give some sense of what's happening – as long as we recognize their flaws.

This column does not necessarily reflect the opinion of Bloomberg LP and its owners.

To contact the author of this story:
Cathy O'Neil at coneil19@bloomberg.net

To contact the editor responsible for this story:
Mark Whitehouse at mwhitehouse1@bloomberg.net