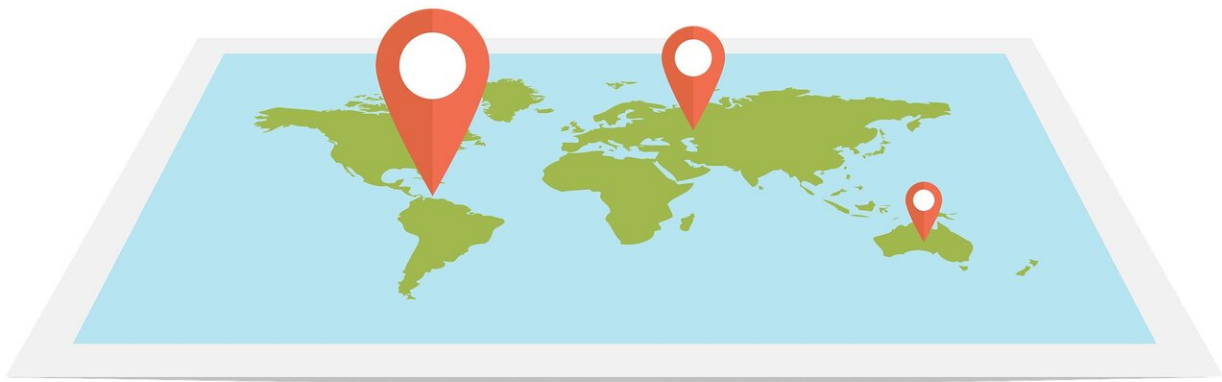# Self-reported COVID-19 symptoms show promise for disease forecasts

April 21 2020

Credit: CC0 Public Domain

Self-reported descriptions of COVID-19-related symptoms, which Carnegie Mellon University researchers are gathering nationwide with the help of Facebook and Google, correlate well with test-confirmed cases of the disease, suggesting self-reports might soon help the researchers in forecasting COVID-19 activity.

Ryan Tibshirani, co-leader of Carnegie Mellon's [Delphi COVID-19 Response Team](link), said millions of responses to CMU surveys by Facebook and Google users are providing the team with real-time estimates of disease activity at the county level for much of the United States.

"I'm very happy with both the Facebook and Google survey results," said Tibshirani, associate professor of statistics and machine learning. "They both have exceeded my expectations."

The [survey results](link), combined with data from additional sources, provide real-time indications of COVID-19 activity not previously available from any other source.

This information will be made publicly available at CMU's COVIDcast web site, [COVIDcast.cmu.edu](link), and Facebook is making the aggregated survey information from its users available at https://dataforgood.fb.com/COVID-symptoms-map/.

CMU launched its COVIDcast site today, featuring estimates of coronavirus activity based on those same surveys from Facebook users. Later this week, the COVIDcast site will debut interactive heat maps of the United States, displaying survey estimates from not only Facebook, but also Google users. The maps also will include anonymized data provided by other partners, including Quidel Corp. and a national health care provider.

Tibshirani said the survey responses combined with other data such as medical claims and medical testing, will enable the CMU team to generate estimates of disease activity that are more reflective of reality than what is now available from positive coronavirus tests alone. Most of the data sources are available on a county level and the researchers say they have good coverage of the 601 U.S. counties with at least 100,000

people.

Within a few weeks, they expect to use these estimates to provide forecasts that will help hospitals, first responders and other health officials anticipate the number of COVID-19 hospitalizations and ICU admits likely to occur in their locales several weeks in advance.

Thus far, CMU is receiving about one million responses per week from Facebook users. Last week, almost 600,000 users of the Google Opinion Rewards and AdMob apps were answering another CMU survey each day.

Using these and other unique data sources, the CMU researchers will monitor changes over time, enabling them to forecast COVID-19 activity several weeks into the future. They also plan to use this information to produce "nowcasts," which are integrated estimates of current disease activity that they expect will be more reflective of reality than are daily compilations of test-confirmed COVID-19 cases.

Roni Rosenfeld, co-leader of the CMU Delphi research group and head of the Machine Learning Department, said relying only on positive test results may not provide a complete picture of disease activity because of limited test capacity, reporting delays and other factors.

For this COVID-19 project, Carnegie Mellon's Delphi research group, which has now grown to include about 30 faculty members, students, and other volunteers, is leveraging years of expertise as the preeminent academic center for forecasting influenza activity nationwide. Last year, the U.S. Centers for Disease Control and Prevention designated the Delphi group as one of two National Centers of Excellence for Influenza Forecasting. At CDC's request, the group has extended and adapted its flu forecasting efforts to encompass COVID-19.

Delphi uses two main approaches to forecasts, both of which have proven effective regarding the flu. One, called Crowdcast, is a "wisdom of the crowds" approach, which bases its predictions on the aggregate judgments of human volunteers who submit weekly estimates. The other uses statistical machine learning to recognize patterns in health care data that relate to past experience.

"This forecasting problem is so complicated that we believe that a diversity of data and approaches is our best weapon," Tibshirani said.

To aid in COVID-19 forecasting, Facebook each day invites some of its U.S. users to voluntarily answer a CMU survey about any COVID-19 symptoms they might be experiencing; the survey is controlled by CMU and individual responses are not shared with Facebook.

Likewise, Google is helping CMU distribute one-question surveys to its users; results also are not shared with Google. Since 2016 Google Health Trends has been providing CMU information about searches that its users perform each day for flu, and more recently for COVID-19-related terms. A major healthcare care provider is sharing anonymized inpatient and outpatient COVID-related counts, and Quidel, a diagnostic test provider, is sharing anonymized national lab test statistics.

Rosenfeld said they hope to bolster their forecasting efforts by adding another five data sources in the next several weeks.

"We're deeply appreciative of the help we are receiving from Facebook, Google and our other partners," Rosenfeld said. "The data they provide is priceless and will give us greater confidence once we are able to begin our forecasts for this deadly disease."

Provided by Carnegie Mellon University