Internet search data can help predict a looming 'twindemic'

April 21 2023, by Joshua Stewart
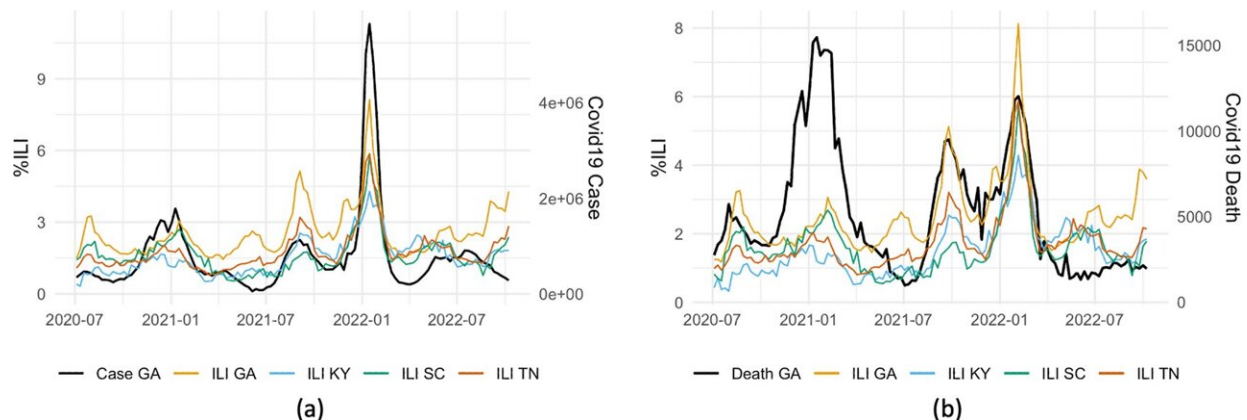

Illustration of Georgia (GA)'s real-time COVID-19 cases/deaths (black) growth in comparision with its own lagged %ILI (yellow) and the lagged %ILI of the neighboring states, from 2020-07-04 to 2022-08-13. a COVID-19 real-time cases in GA (thick black curve) vs. lagged 1 week %ILI in GA (yellow curve) and the neighboring states; (b) COVID-19 real-time deaths in GA (thick black curve) and lagged 3 weeks %ILI in GA (yellow curve) and the neighboring states. Credit: *Communications Medicine* (2023). DOI: 10.1038/s43856-023-00272-2

The most widely used source of medical advice in modern society might be the Google search box.

Enough people turn to the site with searches like "loss of taste" or "how long contagious" that researchers at Georgia Tech can use that data to

accurately predict looming waves of influenza-like illness and COVID-19 infections. Their forecasting models work for the nation overall and for each state, offering a new source of data about potential "twindemics" that could burden health care systems.

The model, developed by Shihao Yang and his team in the H. Milton Stewart School of Industrial and Systems Engineering, is published in the Nature journal *Communications Medicine*.

"Our contribution is that we provide a unique angle for our forecasts: We find that there might be a hack for short-term prediction based on the search data," said Yang, an assistant professor and co-author alongside Georgia Tech Ph.D. student Simin Ma and Shaoyang Ning, assistant professor of statistics at Williams College.

The team's "hack" uses 23 key search queries, like "loss of taste," to look ahead four weeks at the serious cases of both flu-like illnesses and COVID-19. The model uses federal data on hospitalizations and deaths from COVID-19, plus outpatient visits for flu-like illnesses—the official benchmark since it's difficult to separate flu cases from other kinds of viral infections with similar symptoms.

"We provide a unique perspective. Many other people don't really use this data—what you might think about as alternative data," Yang said. "Health care professionals and epidemiologists use surveillance data, survey data, hospital records. Those are the usual realm. As a statistician working in an engineering context, my lens is a bit different on our health care system."

Yang said traditional epidemiological models are much more effective than his approach at forecasting six months or more in the future. Where his team's work shines is in the short-term, offering public health officials and hospitals data to anticipate an imminent surge.

The researchers have presented their research results to the Centers for Disease Control and Prevention.

Yang has been using search data for flu forecasting for years and started to apply that knowledge to the coronavirus pandemic in early 2020. Over time, he sensed the new tools he and others developed for COVID predictions could help with flu modeling and vice versa.

With increasing concern last fall about a "twindemic" of both illnesses, "we realized this seems to be the time to start treating flu and COVID kind of equally," Yang said. "So, the whole idea is, why don't I just put the two diseases together and build a joint model for both."

More information: Simin Ma et al, Joint COVID-19 and influenza-like illness forecasts in the United States using internet search information, *Communications Medicine* (2023). [DOI: 10.1038/s43856-023-00272-2](https://doi.org/10.1038/s43856-023-00272-2)

Provided by Georgia Institute of Technology