

# Can 'digital traces' from internet searches and social media predict outbreaks of COVID-19?

January 19 2023, by Cynthia McCormick Hibbert



A summary of our results at the county level. (A) Graphical representation of the different outcomes observed in our methods: early warnings, synchronous warnings, late warnings, soft warnings, missed outbreaks, warnings with increased activity, and false alarms. (B) Summary of the outbreak onset events. Horizontal bars are colored, from orange to purple, depending on the event class.

(C) False alarms for the Naive, Single Source, and Multiple Source methods. The Multiple Source method produced the lowest amount of false alarms (110). (D) Probability of resurgence  $P(R_t > 1)$  and different events generated by the Naive, Single Source (Google Trends “How long does covid last?” and “side effects of vaccine”), and Multiple Source methods. (E) Earliness of the alarms triggered by each method. Bars represent the number of alarms within the out-of-sample time window. Credit: *Science Advances* (2023). DOI: 10.1126/sciadv.abq0199

Your Google searches and Twitter accounts alert marketers about what items you might like to purchase. But could they also serve as an early warning system when COVID-19 levels are about to take off?

A team of scientists including Northeastern University machine learning expert Mauricio Santillana says internet users' "digital traces" can be adopted to alert [public health officials](#) to sharp increases in COVID-19 at the county level one to six weeks ahead of a major outbreak.

In a paper published Wednesday, Jan. 18, in *Science Advances*, Santillana and other authors say [digital data](#) will help close information gaps left by existing surveillance methods.

Analysis of the data streams will allow policymakers to get a jump on decisions such as whether to reissue masking recommendations or bump up vaccination and boosting campaigns, says Santillana, director of the Machine Intelligence Group for the Betterment of Health and the Environment at the Network Science Institute at Northeastern.

"What we aspire to do is to use the same information that Google or Amazon or any of these big firms use to send ads to you" to inform public health decisions early on in an outbreak, Santillana says.

COVID-19-related digital streams include [internet searches](#) for fever, clinician searches for COVID-19 treatments and Twitter users' comments about being too sick to work, among other things.

The researchers also used machine learning methods that took historic information from outbreaks in 97 U.S. counties from Jan. 1, 2020, and 2022 and combined them to create a single predictive indicator.

"The goal is not necessarily to quantify how many infections there are but to quantify when sharp increases in infections will happen," says Santillana, who participated in the research with scientists from Boston Children's Hospital, Harvard Medical School, Oklahoma State University and other organizations.

Researchers found that the predictive capacity at the state and county levels was roughly similar—the [early warning system](#) deployed at one to six weeks in advance at the county level and four to six weeks at the state level.

The study says the digital data will help fill in vital missing information for the Centers for Disease Control and Prevention, which it says has failed to reliably forecast "rapid changes in the trends of reported cases and hospitalizations."

"When CDC COVID-19 forecasts were shared with the public, they very frequently missed the timing of when outbreaks were starting," Santillana says. He says by the time actual case numbers were tallied, surges were already well under way.

"The next chapter would be for the CDC to say, 'We know that this is an alternative and complementary way to anticipate outbreaks. We will implement it inhouse, and we will have it as an additional tool in our toolbox,'" Santillana says.

"He says the study is part of a new CDC initiative started by President Biden called the Center for Forecasting and Outbreak Analytics within the CDC."

"It is within that effort that we did the work in this paper," published in an open access journal of the American Association of the Advancement of Science, Santillana says.

He says he and his team already had been working with the CDC for three to four years on predicting flu incidence and flu hospitalizations, but he wasn't satisfied with what he considered the CDC's inability to incorporate novel Internet-based sources of information into their prediction systems."

"When COVID hit, they called and said, 'We need all hands on deck. So please do what you can.'"

"I asked if they could be flexible, because my team and I were interested in innovating rather than just continuing to implement the exact same models," Santillana says.

"The model is not perfect," he says.

The counties studied were only a fraction of the 3,006 counties in the United States, according to the paper on using digital traces to build prospective and real-time county-level early warning systems.

"Our internet search-based methods may struggle to perform well in areas with poor literacy rates and limited access to internet resources," the paper says.

The researchers say a possible solution for counties with poor internet access or literacy challenges may be to use state-level early warning

systems to guide county-level decisions around outbreaks.

"When we navigate the internet on our computer or phone it leaves traces," Santillana says.

"Whether we like it or not, the reality is that most companies use this information to increase their profit or their margins," he says.

"Instead, we want to use that information to inform public officials when the next [outbreak](#) will happen."

**More information:** Lucas M. Stolerman et al, Using digital traces to build prospective and real-time county-level early warning systems to anticipate COVID-19 outbreaks in the United States, *Science Advances* (2023). [DOI: 10.1126/sciadv.abq0199](https://doi.org/10.1126/sciadv.abq0199)

Provided by Northeastern University

Citation: Can 'digital traces' from internet searches and social media predict outbreaks of COVID-19? (2023, January 19) retrieved 19 March 2023 from <https://medicalxpress.com/news/2023-01-digital-internet-social-media-outbreaks.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.