Google Searches Can Help Us Find Emerging Covid-19 Outbreaks

They can also reveal symptoms that at first went undetected. I may have found a new one.

By **Seth Stephens-Davidowitz**

Mr. Stephens-Davidowitz is the author of "Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are."

April 5, 2020

Every day, millions of people around the world type their health symptoms into Google. We can use these searches to help detect unknown Covid-19 outbreaks, particularly in parts of the world with poor testing infrastructure.

To see the potential information lying in plain sight in Google data, consider searches for "I can't smell." There is now strong evidence that anosmia, or loss of smell, is a symptom of Covid-19, with some estimates suggesting that 30-60 percent of people with the disease experience this symptom. In the United States, in the week ending this past Saturday, searches for "I can't smell" were highest in New York, New Jersey, Louisiana, and Michigan — four of the states with the highest prevalence of Covid-19. In fact, searches related to loss of smell during this period almost perfectly matched state-level disease prevalence rates, as the accompanying chart shows.

"Loss of Smell"

Google searches for the phrase "loss of smell" align closely with the number of positive cases of coronavirus. The inability to smell could be an early warning sign that someone is infected.

Source: Google | By The New York Times
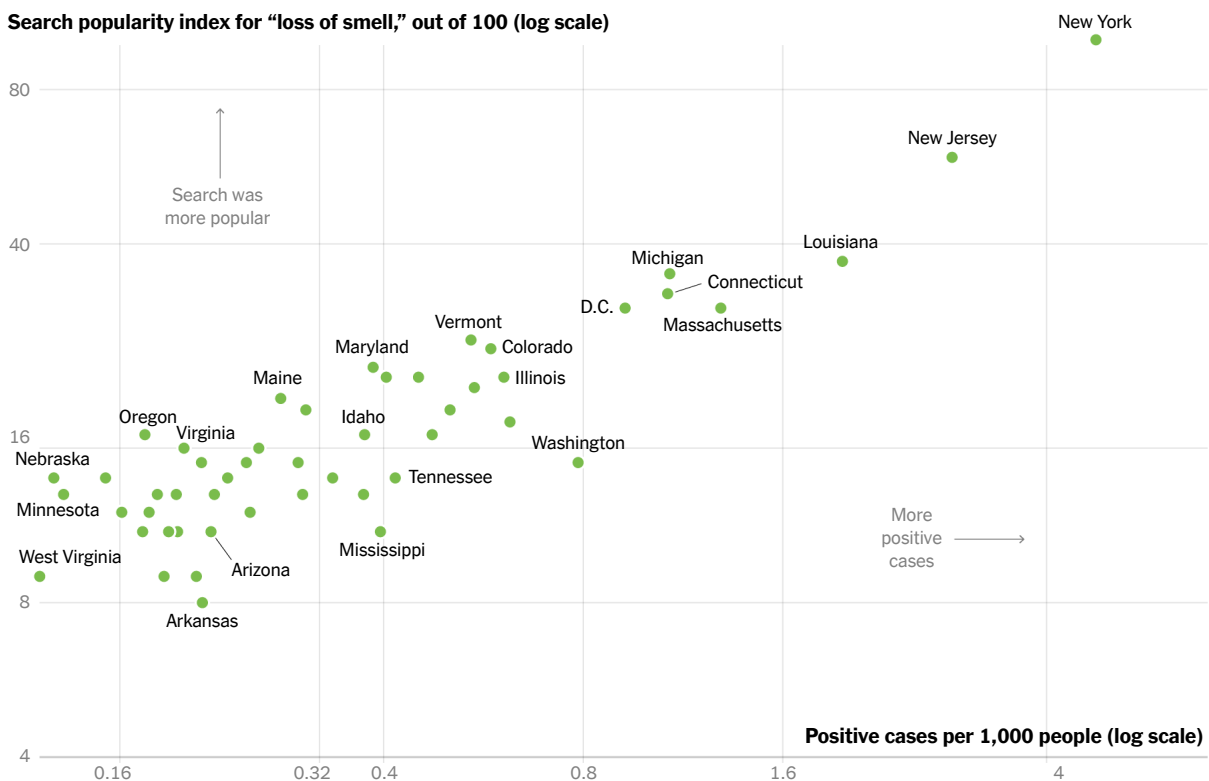
Vasileios Lampos, a computer scientist at University College London, and other researchers have found that a bevy of symptom-related searches — loss of smell as well as fever and shortness of breath — have tracked outbreaks around the world.

Because these searches correlate so strongly with disease prevalence rates in parts of the world with reasonably good testing, we can use these searches to try to find places where many positive cases are likely to have been missed.

Consider Ecuador. The official data says that while Ecuador has among the highest rates of Covid-19 cases per capita in South America, it has a lower case rate than the United States, Canada, Australia, Iran and most of Europe.

At the same time, Ecuadorians are now making more searches related to loss of smell than any other country in the world, once you adjust for total Google searches. Searches for "no puedo oler" ("I can't smell") are some 10 times higher per Google search in Ecuador than they are in Spain, even though Ecuador officially reports more than ten times fewer Covid-19 cases per capita than Spain does. Ecuadorians are also right near the top in searches for fever, chills and diarrhea.

The search data, in other words, suggests that Ecuador may be even more of a Covid-19 epicenter than the official data says. That could help explain recent videos that have been shared on social media of bodies piled up on the street in Guayaquil, a port city in Ecuador.

Cars and trucks carrying coffins lined up outside a cemetery in Guayaquil, Ecuador, on Thursday. Vicente Gaibor del Pino/Reuters

While there does seem to be important information about Covid-19 prevalence in search data, we have to use great care in building models based on this data and learn from past attempts that tried to use this data to measure the geographic spread of different diseases.

In a 2009 paper that was published in Nature, researchers famously showed that Google searches related to the flu had been closely tracking weekly data on influenza rates from the Centers for Disease Control and Prevention. Researchers used these search terms to build a model to try to help detect epidemics before the official data was collated.

Although the model did work initially, it struggled during the 2009 H1N1 flu pandemic. The problem was that flu was in the news so often, many people were searching for flu not because they were feeling symptoms but because they were feeling curiosity or fear. Concern about flu and Google searches about flu were more in the air than actual flu.

Recently, scholars have produced new methods to improve Google-based disease prevalence modeling and helped revive the influenza-tracking project. They have found that it is crucial to key in on the types of searches that are most likely to be reports of symptoms rather than searches related to news.

These tools are being used right now by researchers studying how searches might track Covid-19. Searches like "I can't smell" are particularly useful because the form of the query suggests that someone may have the disease, whereas other queries related to loss of smell may instead suggest curiosity in the topic.

There is another way we can use search data during this pandemic: to better understand symptom patterns of the disease. Our understanding of the progression of symptoms of the disease is still developing. It took until March 20 for widespread reports of the relationship between Covid-19 and loss of smell to surface, even though it now appears to be among the most common symptoms.

There is already some evidence that clues to this symptom were evident earlier in search data. Joshua Gans, a professor at the Rotman School of Management at the University of Toronto, found that searches for "non sento odori" ("I can't smell") were elevated in Italy days before the symptom was reported in the news. Iran also saw an enormous rise in searches related to loss of smell weeks before media reports of the symptom became common.

"Non Sento Odori"

Searches for "non sento odori" ("I can't smell") spiked in Italy as the coronavirus outbreak spread — but before a report was released identifying the possible symptom.

Search popularity index out of 100, three-day moving average
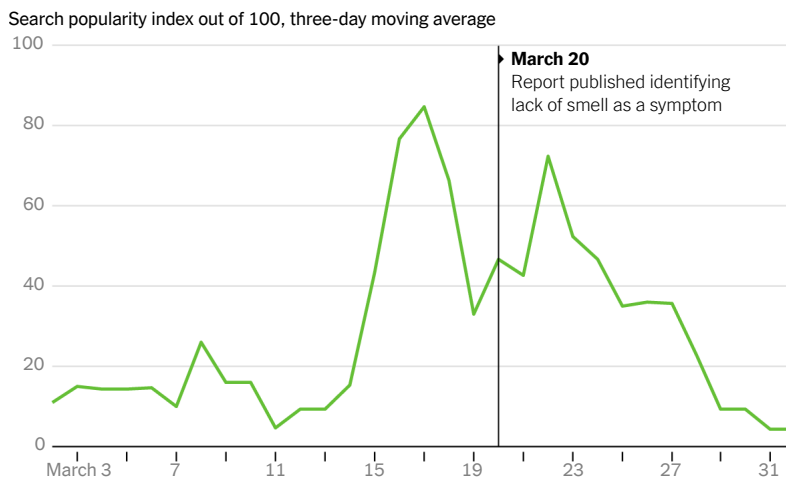
Source: Google | By The New York Times

This would not have been the first time symptom patterns were evident in search query patterns before they were fully recognized by the medical community. In 2016, researchers reported that subtle patterns of searches for symptoms could predict future pancreatic cancer patients. If a person searched for indigestion and later abdominal pain, for example, they were at risk of later searching "just diagnosed with pancreatic cancer." Many of the precise timing patterns of symptoms leading to pancreatic cancer diagnoses were not previously understood.

I have spent the past decade as a data scientist studying how Google searches and other digital data sources can help us measure a range of social and behavioral outcomes. While I am not a medical expert, I was motivated by Professor Gans and others to explore whether Google searches might give clues of symptoms of Covid-19 that have not yet been officially reported. I think I have found a candidate symptom that might come with the disease for at least a small fraction of patients.

First, I downloaded state-level Google search data in the previous week for dozens of symptoms I gathered from medicinenet.com. Next, I measured which searches were most related to a state's disease prevalence rate. In other words, I explored the question of which symptoms are now being searched in unusually high numbers in states with unusually high rates of Covid-19.

The three searches most related to Covid-19 disease rates were not a surprise: loss of smell, fever and chills. The fifth and sixth searches weren't much of a surprise either: nasal congestion and diarrhea, which have also received a lot of attention as Covid-19 symptoms.

However, the fourth-place search *was* a surprise: eye pain, which has not garnered much attention as a possible symptom of the disease. Searches for "my eyes hurt" over the previous week were highest in New York, New Jersey, Connecticut, Louisiana and Michigan. Such searches seem to have risen in the past two weeks almost exclusively in parts of the country that have reached very high Covid-19 rates (although the data is fairly noisy and the rise isn't as large as it is for some other symptoms).

There have been some reports of eye-related issues related to Covid-19. A March 31 paper based on a study of 38 patients in China reported that one third of the patients later tested for ocular irregularities. There have also been recent reports of pink eye in 1-3 percent of Covid-19 patients. Searches related to pink eye do not show nearly as strong a geographic relationship with Covid-19 rates as eye pain, though. In fact, all eye-related complaints except pain that I looked at show little-to-no relationship with Covid-19 rates.

Does the Google search data really mean that eye pain is a symptom of Covid-19? Not necessarily. There may be other reasons that people in these parts of the country are searching for eye pain. However, I tested alternative explanations that people suggested to me, and they did not fit the data. The searches do not seem to be driven by allergies; they are not related to pollen concentrations. Nor do they seem to be driven by people staying at home and staring at screens more; eye pain search rates do not correlate with data from cellphones that have measured recent reductions in movement.

It is hard to imagine that curiosity alone is driving the relationship between eye pain and Covid-19 prevalence rates. Other potential symptoms that have received extensive media attention don't show nearly as strong a statewide relationship with Covid-19 prevalence rates.

There is also some evidence for eye pain as a symptom of Covid-19 from searches in other parts of the world. Notably, searches for eye pain rose above fourfold in Spain between the middle of February and the middle of March and rose about 50 percent in Iran in March. Eye pain searches seemingly haven't risen in Italy, though that data is noisier. (To examine data across the world, I am using Google's topic "eye pain," which groups together many different searches in various languages related to the topic. Since Google reports different random samples of their data for different data requests, I have averaged a number of different samples.)

I think search data offers suggestive evidence that eye pain can be a symptom of the disease. However, it might only affect a small fraction of Covid-19 patients. Overall search volume for eye pain, despite rising substantially in Covid-19 hot spots, remains well below search volume for other symptoms. In New York there are now about one-sixth as many searches related to eye pain as there are searches related to loss of smell.

Nonetheless, doctors and public health officials should probably look closely at the relationship between Covid-19 and eye pain. If nothing else, we need to understand why there is frequently a large uptick in people telling Google that their eyes hurt when known cases of Covid-19 in a location rise to extremely high levels.

More people can study search trends around the world to help us learn about Covid-19. In 2006, Google released Google Trends, a public tool that the research community can use to study anonymous and aggregate search data. That is how I found everything I reported in this piece. It is plausible that important facts about the Covid-19 disease could be found here or in other large data sets by data scientists, medical experts or even amateur data sleuths.

Seth Stephens-Davidowitz (@SethS_D) is an economist, the author of "Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are" and a contributing opinion writer.