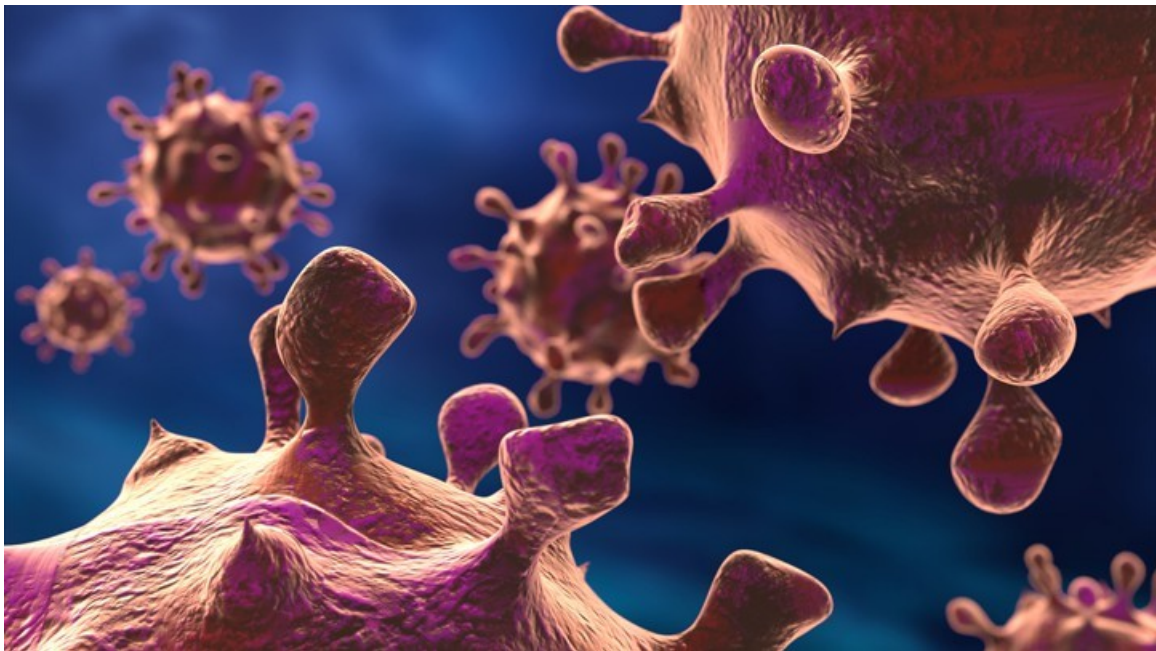

TECHNOLOGY

In Defense of Google Flu Trends

If the system failed, it did so only in the outsize dreams of Big Data acolytes.

ALEXIS C. MADRIGAL MAR 27, 2014



SHUTTERSTOCK/GIOVANNI CANCEMI

In 2008, Google released an experiment called Flu Trends, which attempted to predict the prevalence of the flu from searches that users made for about 40 flu-related queries.

Based on the data up to that point in time, Flu Trends worked really well. The Centers for Disease Control, which had been involved in shaping how it functioned, liked the data that it produced.

"We really are excited about the future of using different technologies, including technology like this, in trying to figure out if there's better ways to do surveillance for outbreaks of influenza or any other diseases in the United States," Joseph Bresee, the chief of the epidemiology and prevention branch in the CDC's influenza division, said at the time.

And so, aside from some misplaced nagging about privacy, the new tool was celebrated in the news media. All the big outlets covered it: CNN, The New York Times, the Wall Street Journal, and many, many more.

Flu Trends fit the golden image of Google, circa 2008: a company that did gee-whiz things just because they were good ideas.

This was the era of Google.org, formed under a guy named *Larry Brilliant*, who said things like, "I envision a kid (in Africa) getting online and finding that there is an outbreak of cholera down the street. I envision someone in Cambodia finding out that there is leprosy across the street." As the global economy collapsed, Google gleamed among the ruins.

This was all years before people started even talking about "big data."

But times have changed. Now, data talk is everywhere, and people are more worried than excited. The NSA looms over every tech discussion.

And Google no longer seems so different from other companies. As Bill Gates put it in an interview with *Bloomberg BusinessWeek*, "Google started out saying they were going to do a broad set of things. They hired Larry Brilliant, and they got fantastic publicity. And then they shut it all down. Now they're just doing their core thing." Larry Brilliant left in 2009. Google.org redirects to Google.com/giving/.

Which brings us back to Google Flu Trends. If you've been watching the headlines this month, you've probably seen that a study published in the journal *Science* took Flu Trends to task for "big data hubris." A team led by Northeastern political scientist David Lazer found that Flu Trends has been missing high, including "100 out of 108 weeks starting with August 2011." They pointed to problems with the opacity of Google's method and the inconsistency of the Google search user interface and algorithms. The paper was titled, "The Parable of Google Flu: Traps in Big Data Analysis."

And the response to it in the media was swift and rough. Here is a list of just a few of the headlines that followed Lazer's methodological fisting:

Google Flu Trends Gets It Wrong Three Years Running

Why Google Flu Is a Failure

A Case of Good Data Gone Bad

Data Fail! How Google Flu Trends Fell Way Short

Google Flu Trends Failure Shows Drawbacks of Big Data

Even the reliable pro-nerd hangout Slashdot headlined their thread on the story, "Google Flu Trends Suggests Limits of Crowdsourcing."

A skim of the headlines and most of the stories would lead you to believe that Google Flu Trends had gone horribly wrong. Here, this thing was nominally supposed to predict future CDC reports—and it wasn't even as good as simple extrapolations from past CDC reports.

How you like them apples, Google/Big Data!

But lurking in the Lazer paper was an interesting fact about the Google Flu Trends data: when you combined it with the CDC's standard monitoring, you actually got a *better* result than either could provide alone. "Greater value can be obtained by combining GFT with other near-real time health data," Lazer wrote. "For example, by combining GFT and lagged CDC data."

If that was true, and the CDC was aware of it, couldn't they simply combine the data on their own and have a better epidemiological understanding of the country? And if that was true, wasn't Flu Trends a success, at least according to the standards laid out in the Nature paper describing it in 2009?

"The system is not designed to be a replacement for traditional surveillance networks or supplant the need for laboratory-based diagnoses or surveillance," Google and CDC authors wrote. They continued, "As with other syndromic surveillance systems, the data are most useful as a means to spur further investigation and collection of direct measures of disease activity."

In other words, Google Flu Trends is not a magical tool that replaces the CDC with an algorithm. But, then, the people who built it never imagined that it was. If it failed, it did so, more than anywhere else, in the popular imagination—and in the wishes of superficial Big Data acolytes.

* * *

When Matt Mohebbi, who spearheaded the creation of Google Flu Trends with Jeremy Ginsberg, started at Google in 2004, his first boss was Peter Norvig, the artificial intelligence legend. "He wrote my textbook in college," Mohebbi told me, "and I debated whether I should bring my textbook in and have him sign it."

It was a heady time. Mohebbi and Ginsberg used their (now all but eliminated) "20-percent time" to see if they could measure disease incidence from search query data. Their initial results were promising, so they pitched the idea up the Google ladder. Eventually, they got sign-off from executive Craig Neville-Manning, now director of New York Engineering at the company, to take three engineers full-time and build out what became Flu Trends.

They immediately got in touch with people at the Centers for Disease Control to find out how to make the tool useful for them. Their early-and-frequent contact with the CDC are "why the system is built the way that it is," Mohebbi said. "The goal was to build a *complementary* signal to other signals."

What that means is this: they didn't want to enmesh their data with the CDC's because then it couldn't act as a separate way of understanding a given epidemiological scenario. In practical terms, if they followed Lazer's advice to combine Flu Trends data with CDC data, it would offer less to the CDC, even if it gave greater predictive value to the Google effort.

"In talking with the various public health officials over the years, we've gotten the feeling that it is very beneficial to see multiple angles on the flu," Mohebbi said.

The other argument against building a model that incorporates CDC data, Mohebbi says, is that public health officials are most interested in when flu trends deviate from a simple "project ahead a few weeks" kind of model. By keeping the Flu Trends signal pure, officials can check to see whether the GFT data reflects a real deviation from the standard trend, or whether it is an artifact of the model.

"We put the data out there so that people can build these more elaborate models," Mohebbi told me, "And in fact, they have."

(The "we" no longer actually includes Mohebbi: he left Google a few years ago and has co-founded a startup called Iodine with

former *Wired* executive editor Thomas Goetz, to help patients make better medication decisions.)

Take this research out of Johns Hopkins, published in early 2013: a team examined how to build a better influenza model, starting with emergency-room clinical data and trying to add anything else that might help including variables such as "GFT, meteorological data (temperature, change in temperature, and relative humidity) and temporal variables (Julian weeks, and seasonality)."

The team found that Flu Trend data "was the only source of external information to provide statistically significant forecast improvements over the base model." That is to say, it played the precise role that it was designed for: providing a complementary signal.

All of this doesn't explain or excuse why Flu Trends isn't working as well as it did in its initial runs. Perhaps the model overfit flu data. Or perhaps Google adding more suggestions for users threw off the baseline. Or perhaps the underlying media ecosystems or user search behavior are changing more quickly than anticipated.

But let's be reasonable here: Mohebbi, Ginsberg, and Google created a thing that helps the CDC and other epidemiologists. It's not perfect, and the ways that it is imperfect are instructive—thus, Lazer using it as a parable of big data techniques' problems.

But it is still useful to the people who matter.

I asked Mohebbi whether he drew a different lesson from the Google Flu Trends "parable."

"There is great value that comes with triangulating these [flu tracking] systems," he said. "There isn't a ground truth for what exists. They are all proxies for flu illness or influenza-like illness, and there is great power that comes from combining these signals."

He concluded: "There's huge promise with these techniques, but you have to understand how they should be used."

I also think that the Google Flu Trends story is a parable, but it goes more like this:

New technology comes along. The hype that surrounds it exceeds that which its creators intended. The technology fails to live up to that false hope and is therefore declared a failure in the court of public opinion.

Luckily, that's not the only arena that matters. Researchers both in and outside epidemiology have found Google Flu Trends and its methods useful and relevant. The initial *Nature* paper describing the experiment now has over 1,000 citations in many different fields.

We want to hear what you think about this article. Submit a letter to the editor or write to letters@theatlantic.com.
