

**Steven Salzberg** Contributor*Fighting Pseudoscience*

Opinions expressed by Forbes Contributors are their own.

PHARMA & HEALTHCARE 3/23/2014 @ 9:00AM | 54,129 views

Why Google Flu Is A Failure

It seemed like such a good idea at the time.



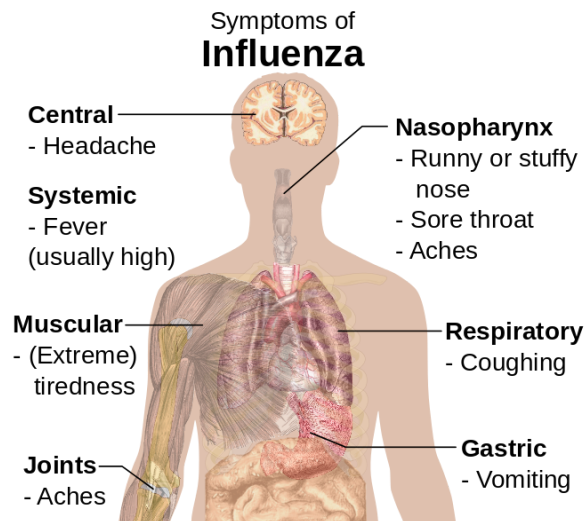
People with the flu (the influenza virus, that is) will probably go online to find out how to treat it, or to search for other information about the flu. So Google decided to track such behavior, hoping it might be able to predict flu outbreaks even faster than traditional health authorities such as the Centers for Disease Control (CDC).

Instead, as the authors of [a new article in Science](#) explain, we got “big data hubris.”

David Lazer and colleagues explain that:

“Big data hubris” is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

The folks at Google figured that, with all their massive data, they could outsmart anyone.



Symptoms of influenza. Model: Mikael Häggström. (Photo credit: Wikipedia)



The problem is that most people don't know what "the flu" is, and relying on Google searches by people who may be utterly ignorant about the flu does not produce useful information. Or to put it another way, a huge collection of misinformation cannot produce a small gem of true information. Like it or not, a big pile of dreck can only produce more dreck. [GIGO](#).

Google's scientists [first announced Google Flu](#) in a *Nature* article in 2009. With what now seems to be a textbook definition of hubris, they wrote:

“...we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day.”

They obtained this remarkable accuracy entirely from analyzing Google searches. Impressive – if true.

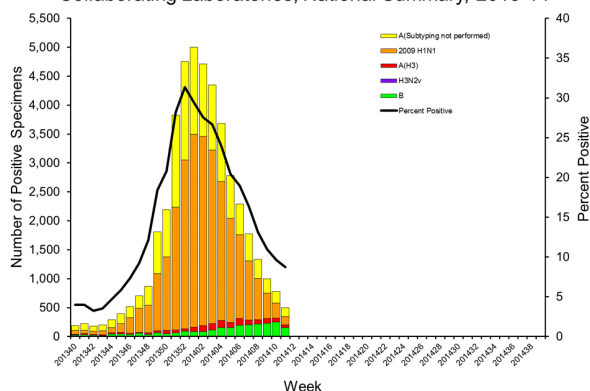
Ironically, just a few months after announcing Google Flu, the world was hit with the 2009 swine flu pandemic, caused by a novel strain of H1N1 influenza. Google Flu missed it.

The failures have continued. As [Lazer et al. show](#) in their *Science* study, Google Flu was wrong for 100 out of 108 weeks since August 2011.

One problem is that Google's scientists have never revealed what search terms they actually use to track the flu. A paper [they published in 2011](#) declares that Google Flu does a great job. The [official Google blog](#) last October makes it appear that they do an almost perfect job predicting the flu for *previous* years.

Haven't these guys been paying attention? It's easy to predict the past. Does anyone remember the University of Colorado professors who had a model that correctly predicted every election since 1980? In August 2012, they confidently announced that their model showed Mitt Romney winning in a landslide. Hmm.

Influenza Positive Tests Reported to CDC by U.S. WHO/NREVSS
Collaborating Laboratories, National Summary, 2013-14



Current flu activity, dominated by H1N1. Source: CDC.

A bigger problem with Google Flu, though, is that most people who think they have “the flu” do not. The vast majority of doctors’ office visits for flu-like symptoms turn out to be other viruses. CDC tracks these visits under “influenza-like illness” because so many turn out to be something else. To illustrate, [the CDC reports](#) that in the most recent week for which data is available, *only* 8.8% of specimens tested positive for influenza.

When 80-90% of people visiting the doctor for “flu” don’t really have it, you can hardly expect their internet searches to be a reliable source of information.

Google Flu [is still there](#), and you can still look at its predictions, even though we know they are wrong. I recommend [the CDC website](#) instead, which is based on actual data about the influenza virus collected from actual patients. Big data can be great, but not when it’s bad data.

RECOMMENDED BY FORBES

[The Cloud 100](#)

[Yes, You Can Leave A Job Interview -- Here's How](#)

This article is available online at: <http://onforbes.es/1m0DTDs>

2016 Forbes.com LLC™ All Rights Reserved