A little over a year ago, a damning report published in the journal *Nature* (1) sent a ripple through the public health and social science communities: in a season of unusually high flu outbreak, one of the world’s newest biosurveillance systems was inaccurately predicting flu trends.

Google Flu Trends (GFT),3 launched in 2008, is an attempt by Google to predict, in real time, the prevalence of influenza-like infection (2).

In combination with computer modeling, the company mines data records for flu-related keyword searches, producing a map of on-the-ground flu activity. GFT predicts trends at the local and state levels, and nationally in at least 29 countries around the world (3).

“The GFT effort is a prominent one in the field,” says Dr. David Lazer, a professor of political science and computer and information science at Northeastern University, in an e-mail to *Clinical Chemistry*. As his professional title would indicate, Lazer is no stranger to big data and has been a proponent for big data analysis in public health.

The *Nature* report highlights that in the peak of the 2012–2013 US flu outbreak, GFT’s national “peak of flu” was almost double that of the CDC. (The report also points out that in other countries, GFT was hitting projected targets.)

Lazer, an early supporter of GFT, admits he had not closely examined the product.

“It then was widely reported that it had missed by a huge margin last winter, which begged the question: why?” he says.

This past March, Lazer and a team of colleagues from Northeastern and Harvard universities published their answer to that question in the journal *Science* (4).

What Is It?

In looking at GFT trends, Lazer and colleagues write that since well before 2013, “GFT had been persistently overestimating flu prevalence.”

For example, the team found that for the 2011–2012 flu season, GFT “missed high” compared to the CDC data “for 100 out of 108 weeks.” And that was after GFT was tweaked in 2009, when it had “completely missed the nonseasonal 2009 influenza A-H1N1 pandemic,” making it, the authors write, “part flu detector, part winter detector” (Fig. 1).

In their explanation of how GFT missed the mark, Lazer and his coauthors home in on 2 points: what they describe as big data hubris and algorithm dynamics.

The former, the authors explain, is the “assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.”

“The comparative value of the algorithm as a stand-alone flu monitor is questionable,” write the authors.

They then go on to demonstrate that GFT would improve its analysis by combining its data with others; for example, the authors combine recalibrated GFT data with lagged CDC data (typically by 2 weeks), and both tools saw performance increase.

As for algorithm dynamics, this is the idea that software engineers are constantly adapting the Google search algorithms, and users are constantly changing their behavior, both of which combine to affect GFT’s tracking.

“In improving its service to customers, Google is also changing the data-generating process,” the authors write. “Because GFT uses the relative prevalence of search terms in its model, improvements in the search algorithm can adversely affect GFT’s estimates.”
Obviously, the authors concede, Google needs to continuously improve its search algorithms, but if they are serious about GFT, they need to find a way to marry that with their flu prediction models.

"Companies need to decide—are you a cute party trick, or the real thing?” says Dr. Kenneth Mandl, a professor at Harvard Medical School and the Boston Children’s Hospital chair in biomedical informatics and population health, in regard to companies that want to enter the public health sphere.

Mandl was not involved in the Science report, but in a phone interview with Clinical Chemistry says he believes the authors’ critique of Google is appropriate, “as it was in part against Google for not appropriately tweaking in the face of new data.”

“I praise Google for innovative use of their enormous data resources,” says Mandl, “but their model was overfit to early data and did not evolve, so it became out of date as the underlying data changed.”

Every year, there are different strains, different patterns, and different impacts, not to mention pandemics.

“It takes a lot of work to keep up with models,” says Mandl. “And validation against external sources is important.”

If you’re the fun widget, says Mandl, whether you’re Google or some other big data miner like Facebook, Apple, or Twitter, then you should advertise yourself as such; in fact, it’s been suggested that perhaps the reason people are reacting so vehemently to the as of now disappointing GFT experiment is because Google seemed to suggest that their flu prediction system would replace all others, including the CDC. (Though in their original report, the GFT creators do write, “This system is not designed to be a replacement for traditional surveillance networks or supplant the need for laboratory-based diagnoses and surveillance” (2).)
But if such big data companies would like to seriously join the biosurveillance and diagnostics community—and the momentum feels like many think they should—there are steps.

For starters, write the Science authors, transparency is key.

“Even if one had access to all of Google’s data, it would be impossible to replicate the analyses of the original paper from the information provided regarding the analysis,” write the authors of the report in Science.

“I take both clinical and public health apps seriously,” says Mandl, “and if Google produces diagnostic or surveillance apps, they should be subject to community review of the detailed methods.”

Moreover, Google has come under fire for not sharing the 45 flu-related search terms it uses in its analysis, the secret sauce in their analytic kitchen.

“Science is a cumulative endeavor and to stand on the shoulders of giants requires that scientists be able to continually assess work on which they are building,” write the authors of the Science report. “Accumulation of knowledge requires fuel in the form of data.”

“If some element of the data were made available to external researchers, we would have many high-performing versions of GFT,” writes Lazer in an e-mail to Clinical Chemistry. Though he notes, “A tricky question is how transparent you can make the data without creating privacy issues or threaten[ing] proprietary interests of Google.”

And that is no small side-note.

“This is something we really have to address,” says Mandl. “Managing large datasets, and getting many eyes on them, is difficult to do while maintaining privacy.”

Despite some obstacles, and their critique of GFT thus far, the Science team, as well as Mandl and others, are nothing less than optimistic about the potential for big data not only with broad public health trends, but point-of-care diagnostics.

“Big data can provide context around a diagnosis,” says Mandl, who demonstrated this idea in 2011 (5).

For his study, Mandl provided physicians with local biosurveillance of incidence of group A streptococcal (GAS) pharyngitis, the most common bacterial form of strep. In their report, Mandl’s team demonstrates that using biosurveillance data improves “prediction of positivity for GAS pharyngitis in patients presenting with pharyngitis across all clinical risk categories.”

Lazer says he’s skeptical that big data would be used for any “serious” diagnoses, but that it might be applicable for non–life-threatening illness, in the same way we see such things done now.

“I have a stomach bug, see the doctor, who says just take pepto, and then my wife gets a stomach bug, we’d probably just get another bottle of pepto,” says Lazer. “More generally, however, whatever ails my friend is simply more diagnostic information.”

Of big data married to public health, Lazer and his coauthors write, “We contend these are the most exciting frontiers in studying human behavior.”

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References

bts.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/
(Accessed May 2014).

DOI: 10.1373/clinchem.2014.226845