

Where are STDs rampant? Google wants to help researchers find out

December 16 2015, by Mary Chris Jaklevic, Kaiser Health News

With sexually transmitted diseases on the rise, researchers at the University of Illinois at Chicago think they might have a powerful new weapon to fight their spread: Google searches.

The nation's leading search engine has quietly begun giving researchers access to its [data](#) troves to develop analytical models for tracking infectious diseases in real time or close to it. UIC is one of at least four academic institutions that have received access so far, along with the U.S. Centers for Disease Control and Prevention.

Researchers can mine Google data to identify searched phrases that spiked during previous upticks in a particular disease. Then, they measure the frequency of those searches in real time to estimate the number of emerging cases. For instance, a jump in gonorrhea might coincide with more people searching "painful urination" or other symptoms.

"If this works, it could revolutionize STD surveillance," said Supriya Mehta, an associate professor of epidemiology at the UIC School of Public Health.

Search trends can be broken down by city and state, weighted according to their significance and combined with other data sources to give a snapshot of where disease is spreading well before public health agencies report the number of verified cases.

"We're hoping for a bit of creativity to flourish around this," Christian Stefansen, Google disease trends senior engineer, said during a visit to UIC last month, where he spoke to about 100 people about lessons Google learned in its attempts to mine data for public health. "There's no shortage of communicable diseases, sadly."

Sexually transmitted diseases are a growing threat, worsened by the progress of antibiotic-resistant strains, according to the CDC. The agency reported in November that STDs, including chlamydia, gonorrhea and syphilis, all increased in 2014, with chlamydia reaching a record of more than 1.4 million new cases. Diagnoses are highest in 15- to 24-year-olds, an age group where technology use also is high.

Public health advocates have long salivated over the idea of using Internet searches to track all sorts of diseases but were limited to the publicly available Google Trends tool. It restricts the number of phrases that can be tracked and does not report searches that fall below certain undisclosed volume thresholds.

Google invited infectious disease researchers to apply for unrestricted access to search data in August as it disbanded its own real-time tracking tool, Flu Trends. Launched in 2008, Flu Trends broke ground but persistently overpredicted cases, and Google came under fire from some researchers for not disclosing its methodology. According to a paper published in *Science* by some independent researchers, Flu Trends stumbled because it used search terms that correlated with flu season but not actual cases of the flu and failed to adjust after Google introduced "search suggest" and other features to guide users to information.

Google is the most commonly used search engine in the U.S., with a 63.9 percent market share in October, according to comScore, a Reston, Va.-based analytics company.

Google searches can be tracked by city, providing more refined data than the national and multi-state data reported by the CDC. "It's a phenomenal data feed to work with, and there's a lot that can be done with it from a research standpoint," said Jeffrey Shaman, an associate professor in environmental health sciences at Columbia University's Mailman School of Public Health, which was given access to the data.

But no matter how great it is, some researchers say they can't rely on Google alone. Take flu, which is furthest along of any real-time disease-tracking effort, with at least nine teams working with the CDC on 12 forecasting models for the current season. This fall Boston Children's Hospital and Harvard Medical School launched HealthMap FluCast, a tool that gave one- and two-week predictions by incorporating Google searches with the CDC's weekly surveillance reports; electronic medical records from athenahealth; and Flu Near You, a website of patient-reported data. They just launched HealthMap Flu Trends, the site where they will be tracking the flu this season.

In a recent paper FluCast architects say with multiple data sources they produce "more accurate and robust real-time flu predictions than any other existing system." Co-founder John Brownstein said in an interview that FluCast will eventually add data from Twitter, though it's "taking time to get the data in order."

While flu patients may find it therapeutic to tweet about their high fevers, pounding headaches and extreme exhaustion, people who suspect they have a sexually transmitted illness are unlikely to vent about their symptoms via social media.

"In no way shape or form is someone going to tweet, 'I have bumps on my vulva. Do you think it's an STI?'" said Amy Johnson, a UIC Ph.D. candidate who's been studying the feasibility of using search data for tracking sexually transmitted infections.

Mehta agreed: "Because STDs are so stigmatized and personal, Twitter is not going to work for that."

Robust STD tracking systems might incorporate additional search engines such as Yahoo and Bing as well as weekly surveillance reports from local health departments, Johnson said.

Overreliance on one source is particularly risky if it's a private company such as Google, which could remove access at any time. Even the CDC is hedging its bets; Matthew Biggerstaff, an epidemiologist who leads flu tracking efforts there, said the national health agency is exploring whether it can measure visits to its own website as a reliable disease indicator "so we have something that's more of a public data set."

And it remains to be seen just how real-time data could be used by public health agencies and providers. In the coming months, the CDC will be asking state and local health departments what type of [flu](#) data they want - real-time versus three-month forecasting, for example - and how they would use it, Biggerstaff said.

"Producing it and showing that it works is different than operationalizing it," Biggerstaff said. "It's still new in terms of incorporating it into a [public health](#) data stream."

Then, there's the issue of public trust. Researchers emphasize that no one's privacy will be violated. Even with their unprecedented data access, researchers will not be able to tell who performs a query, what their sex or ethnicity is, or even what neighborhood that person lives in, Johnson said.

"I'm not going to knock on their door and tell their wife or their husband they have a sexually transmitted infection," Johnson said. "It's important for people on the individual level to know it's about community health."

©2015 Kaiser Health News
Distributed by Tribune Content Agency, LLC.

Citation: Where are STDs rampant? Google wants to help researchers find out (2015, December 16) retrieved 26 April 2024 from <https://medicalxpress.com/news/2015-12-stds-rampant-google.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.