



A better way to track flu

By Gale Pryor | February 9, 2018

January's two-day shutdown of the federal government – including influenza tracking by CDC in the midst of the most severe season in recent years – underscored the need for backup systems to track infectious diseases.

But what may be needed is not backup systems, but better systems, says John Brownstein, epidemiologist at Harvard Medical School and Chief Innovation Officer at Boston Children's Hospital.

Brownstein, Mauricio Santillana, assistant professor at Harvard Medical School's Computational Health Informatics Program, and a team of researchers looked at the impact of layering internet-based data to track the spread of the flu. The results of their study were published in the *Journal of Medical Internet Research* in January 2018.

Traditional flu tracking relies on public health laboratories reporting specimens that test positive for influenza to the CDC – requiring four to five days for results, and up to three weeks before results are reported to the public. Even when combined with reporting of influenza-like illnesses (ILI) from a network of physicians, the system falls short of real-time reporting and city-level “spatial resolution,” or geographic tracking beyond regional trends.

To test an alternate system, the Harvard team layered internet-based data from Google, Twitter, and Flu Near You, a crowd-sourced platform, with data from athenahealth's cloud-based electronic health record (EHR).

“You're going to have different data streams telling you different things,” says Brownstein. “But if you can bring together many different layers of information –

interactions with clinical visits, discussions about flu on social media, people's engagement in a crowd-sourcing tool – that is going to give you the most robust picture of a flu epidemic. And in turn, it is going to give you earlier valid signals that you can do something about.”

Early Google efforts

Turning to internet-based data sets to predict flu activity has been tried before.

In 2008, Google Flu Trends attempted to track the flu by mining flu-related search terms – and for a few years that worked, matching the CDC's data. But in 2013, the year of the H3N2 “swine flu,” Google overestimated the peak by 140 percent. Media coverage of the pandemic had triggered a flood of flu-related searches by people who were not ill, “confounding” the algorithm, as Nature reported. The initiative was shut down in 2015.

The other behemoth of internet-based data, Twitter, has also shown promise and pitfalls. Real-time analyses of Twitter chatter referencing influenza –#flu, #fever, #sickasadog – can be robust due to the volume of activity, an estimated 500 million tweets a day, and the ability to geolocate tweets. But, says Brownstein, a lot of noise surrounds those signals as flu-related tweets rise and fall with trending topics.

EHRs offer another source of tracking diseases in real time: Cloud-based systems offer near-real-time visibility of patient visits coded for ILI, and the data can be cut by age, gender, and zip code of provider. And that enables tracking to the level of a city.

Matching ‘the gold standard’

In the Harvard study, cloud-based EHR data was layered with data derived from social media platform to track influenza over four flu seasons in Boston. The combined diverse data sets both “nowcasted” the flu for the city – and forecast its prevalence by a week. The results matched the accuracy of the Boston Public Health Commission (BPHC), “the gold standard,” says Santillana, for “ground truth,” or actual diagnoses of the flu in a specific community.

Crossing data sets from social media platforms with EHR

data was the key. “It's been more challenging historically to work with EHR data because of the fact that the data sets are highly siloed. Even if many hospitals and clinical settings are on the same EHR platform, that data isn't linked in any real way that you can make real broad analyses of the data,” says Brownstein.

“But what was unique about cloud-based EHR data is that we could in rapid ways extract outcomes that allowed us to look at how the flu season evolves in real time. And I think that is very exciting.”

“As care digitizes,” says Josh Gray, vice president of research at athenahealth, “there are cases where industry has access data sets that are more current and more granular than what academics have traditionally had access to. That said, academics can be much more detailed and rigorous in analyzing some of these data sets. Industry partnerships with academia are generating useful insights that would not be possible with either working in isolation.”

Gale Pryor is senior editor of athenaInsight.

athenainsight

A daily news hub reporting from the heart of the health care internet, with access to a comprehensive data set of health care transactions from athenahealth's nationwide network. We equip leaders with actionable insight and inspiration for making health care work as it should.

Stay in the know

Sign up for weekly data and news:
insight.athenahealth.com/newsletter