

'Ensemble' modeling could lead to better flu forecasts

October 29 2015

By combining data from a variety of non-traditional sources, a research team led by computational epidemiologists at Boston Children's Hospital has developed predictive models of flu-like activity that provide robust real-time estimates (aka "now-casts") of flu activity and accurate forecasts of flu-like illness levels up to three weeks into the future. The team's findings—published in the journal *PLOS Computational Biology*—show that their approach, called ensemble modeling, results in predictions that are more robust than those generated from any one data source alone, and which rival in real time the accuracy of the CDC's retrospective flu reporting.

"We've focused for many years on using individual data sources for tracking a range of diseases," said study senior author John Brownstein, PhD, Boston Children's chief innovation officer and co-founder of the disease tracking site HealthMap. "This represents the next logical step—combining data in a new way where the whole is more valuable than the sum of its parts.

"Weather forecasting is an established discipline and has become engrained in society," he added. "We think the time is ripe for the same to happen with disease forecasting."

While the CDC closely monitors seasonal flu-like illness activity across the U.S., the data reports it generates and distributes to clinicians and public health authorities is historically one-to-two weeks out of date. As accurate predictions could help guide hospitals and health systems in

allocating resources for flu care, many groups have attempted to create models that could provide accurate real-time snapshots of current and predictions of impending flu activity. The most famous of these attempts is probably Google Flu Trends (GFT), launched in 2008 but was decommissioned in 2015.

"There are many data sources and models that can be used to predict flu-like symptoms in the population," said study lead author Mauricio Santillana, PhD, of Boston Children's Computational Health Informatics Program and the Harvard John A. Paulson School of Engineering and Applied Sciences. "But our question was, if we have many models each predicting flu activity, do we gain anything by combining them?"

Santillana and Brownstein's team started with four separate now-casting models of flu-like illness activity, each fed aggregated, anonymized, national-level data from one of four sources: a) search data from Google; b) Twitter data; c) near-real time clinical data from electronic health record (EHR) manager athenahealth; and d) crowd-sourced flu data from Flu Near You, a participatory surveillance system developed by HealthMap. In an approach similar to that used by weather forecasters to predict hurricane tracks, the team then used machine-learning techniques to generate a set of "ensemble" models that incorporated the results produced by the other four single-source models.

To determine their ensemble models' accuracy and robustness, Santillana and Brownstein's team compared their results to those of each of the four real-time source models, as well as both CDC's historical flu-like illness reports and GFT-based now-casts from the 2013-14 and 2014-15 flu seasons. The ensemble models not only outperformed their four real-time source models, but when compared to CDC's historical flu-like illness reports, generated better forecasts of both the timing and the magnitude of flu-like illness activity at each time horizon measured ("this week," "next week," "in two weeks") than models that rely on

historical information only.

The ensemble predictions also accurately tracked CDC's reports of actual flu activity, with near perfect correlation (0.99 Pearson correlation) for real time estimates and slightly smaller correlation (0.90 Pearson correlation) at the two-week time horizon.

Thus, Santillana points out, the answer to his question is yes. "If we combine multiple data sources, we get a stronger, more robust, more accurate prediction of flu activity."

One of the keys to the [model](#)'s success, he added, is the inclusion of social media and EHR data. "People sometimes wonder if the information that we are getting from social media or EHRs is really valuable, and we could get away with building models based on historical data. But we found that the data sources we had access to provided us with information that was better than just looking at historical patterns."

The researcher team hopes to increase the models' geographic resolution—right now, it only predicts [flu activity](#) on a national scale—as well as extend the models' capabilities to track other diseases where multiple data sources are available (e.g., dengue), and disease activity in other nations. They also hope to produce a publicly available flu prediction tool based on their models.

"What have people in informatics, medicine and public health dreamed of for years? The ability to leverage all manner of data—historic, social, EHR, and so on—to create a learning health system," Brownstein said. "With this approach, we think we've taken a big step in that direction. Our job now is to see if we can refine and expand upon it, and apply it in ways that can benefit as many people as possible."

More information: *PLOS Computational Biology*:

[journals.plos.org/ploscompbiol... journal.pcbi.1004513](https://journals.plos.org/ploscompbiol/article/doi/10.1371/journal.pcbi.1004513)

Provided by Public Library of Science

Citation: 'Ensemble' modeling could lead to better flu forecasts (2015, October 29) retrieved 20 April 2024 from <https://medicalxpress.com/news/2015-10-ensemble-flu.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.