

## Detecting Overlapping Outbreaks of Influenza

John M. Aronis\*<sup>1</sup>, Nicholas E. Millett<sup>1</sup>, Michael M. Wagner<sup>1</sup>, Fuchiang Tsui<sup>1</sup>, Ye Ye<sup>1</sup>, Jeffrey Ferraro<sup>2,3</sup>, Peter J. Haug<sup>2,3</sup> and Gregory F. Cooper<sup>1</sup>

<sup>1</sup>Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA; <sup>2</sup>University of Utah, Salt Lake City, UT, USA;

<sup>3</sup>Intermountain Healthcare, Salt Lake City, UT, USA

### Introduction

Influenza is a contagious disease that causes epidemics in many parts of the world. The World Health Organization estimates that influenza causes three to five million severe illnesses each year and 250,000-500,000 deaths [1]. Predicting and characterizing outbreaks of influenza is an important public health problem and significant progress has been made in predicting single outbreaks. However, multiple temporally overlapping outbreaks are also common. These may be caused by different subtypes or outbreaks in multiple demographic groups. We describe our *Multiple Outbreak Detection System* (MODS) and its performance on two actual outbreaks. This work extends previous work by our group [2,3,4] by using model-averaging and a new method to estimate non-influenza influenza-like illness (NI-ILI). We also apply MODS to a real dataset with a double outbreak.

### Methods

MODS is part of a framework for disease surveillance developed by our group. In this framework, a natural language processing system extracts symptoms from emergency department patient-care reports. These features are combined with laboratory results and passed to a case detection system that infers a probability distribution over the diseases each patient may have. These diseases include influenza, NI-ILI, and other (appendicitis, trauma, etc.). This distribution is expressed in terms of the likelihoods of the patients' data. These are given to MODS which searches a space of multiple outbreak models, computes the likelihood of each model, and calculates the expected number of influenza cases day-by-day. This work differs from past work in three important ways. First, we address the problem of detecting and characterizing multiple, overlapping outbreaks. Second, we do not rely on simple counts, but use likelihoods given evidence in the free-text portion of patient-care reports as well as laboratory findings. Third, we explicitly account for non-influenza influenza-like illnesses. This is important because some forms of influenza-like illness (such as respiratory syncytial virus) are contagious and exhibit outbreak activity. This research was approved by the University of Pittsburgh and Intermountain Healthcare IRBs.

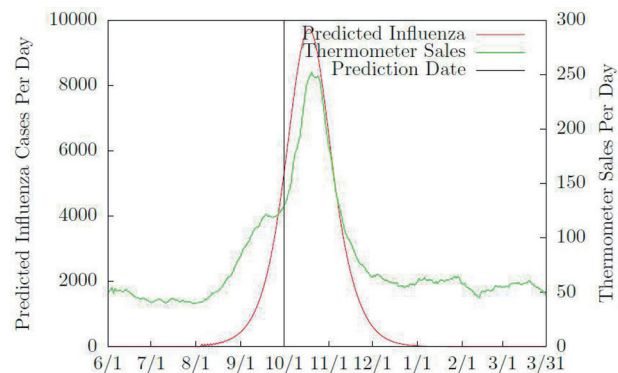
### Results

We conducted a set of experiments with simulated outbreaks. MODS is able to detect a single outbreak six to eight weeks before the peak. It is also able to recognize a second outbreak approximately halfway between peaks for simulated double outbreaks. We conducted experiments using real outbreaks and compared our results to thermometer sales [5]. Using data from Allegheny County Pennsylvania for the 2009-2010 influenza season, on September 1 MODS predicted an outbreak with a peak on October 5. The thermometer peak was October 21. The figure "Prediction on October 1 for Allegheny County" compares MODS' prediction on October 1 to thermometer sales. Using data from Salt Lake City Utah for the 2010-2011 influenza season, on November 1 MODS predicted an outbreak with peak on December 7. The first thermometer peak was December 29. On January 20 MODS predicted a second outbreak with peak on February 9. The second thermometer peak was March

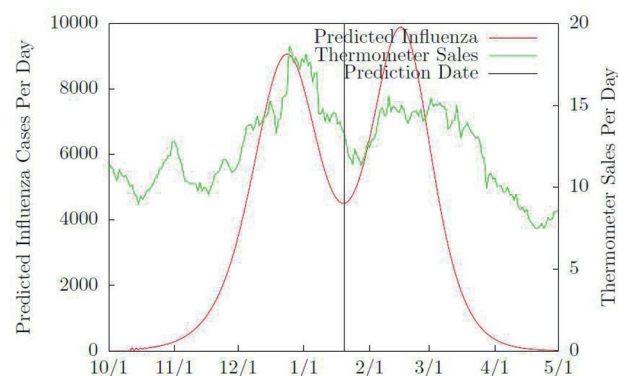
5. The figure "Prediction on January 20 for Salt Lake City" compares MODS' prediction on January 20 to thermometer sales.

### Conclusions

We have built a *Multiple Outbreak Detection System* that can detect and characterize overlapping outbreaks of influenza. Although the system currently predicts outbreaks of influenza, it is built on a general Bayesian framework that can be extended to other diseases. Future work includes incorporating multiple forms of evidence, modeling other known contagious diseases, and detecting outbreaks of new previously unknown diseases.



Prediction on October 1 for Allegheny County 2009-2010



Prediction on January 20 for Salt Lake City 2010-2011

### Keywords

influenza; outbreak; detection; bayesian

### Acknowledgments

This work was supported by NIH grant R01 LM011370 on Probabilistic Disease Surveillance. John Aronis was supported by the National Library of Medicine Training Grant T15LM007059 to the University of Pittsburgh.



## ISDS 2016 Conference Abstracts

### References

1. World Health Organization, Influenza fact sheet, 2003
2. Wagner M, Tsui F-C, Cooper G, Espino J, Levander J, Villamarin R. Probabilistic, decision-theoretic disease surveillance and control. *Online Journal of Public Health Informatics*, 2011, Volume 3, Number 3.
3. Cooper G, Villamarin R, Tsui F, Millett N, Espino J, Wagner M. A method for detecting and characterizing outbreaks of infectious disease from clinical reports. *Journal of Biomedical Informatics*, 2015, Volume 53, 15-26.
4. Aronis J, Millett N, Wagner M, Tsui F, Ye Y, Cooper G. A method for detecting and characterizing multiple outbreaks of infectious diseases. *Online Journal of Public Health Informatics*, 2016, Volume 8, Number 1.
5. Villamarin R, Cooper G, Wagner M, Tsui F, Espino J. A method for estimating from thermometer sales the incidence of diseases that are symptomatically similar to influenza. *Journal of Biomedical Informatics*, 2013, Volume 45, 444-457.

**\*John M. Aronis**

E-mail: [jma18@pitt.edu](mailto:jma18@pitt.edu)

