Emerging Health Threats Journal 2011, **4**:s57. **doi:** 10.3134/ehtj.10.057

www.eht-journal.org

ABSTRACT

Estimating the incidence of influenza cases that present to emergency departments

R Villamarín, G Cooper, F-C Tsui, M Wagner, and J Espino

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: ricardov@pitt.edu

Objective
We developed a model that predicts the incidence of influenza (flu) cases that present to Emergency Departments (ED) in a given region based on thermometer sales (TS).

Introduction
Our laboratory previously established the value of over-the-counter (OTC) sales data for the early detection of disease outbreaks.[1] We found that TS increased significantly and early during influenza season. Recently, the 2009 H1N1 outbreak has highlighted the need for developing methods that not only detect an outbreak but also estimate incidence so that public-health decision makers can allocate appropriate resources in response to an outbreak. Although a few studies.[2] have tried to estimate the H1N1 incidence in the 2009 outbreak, these were done months afterward and were based on data that are either not easy to collect or not available in a timely fashion (for example, surveys or confirmed laboratory cases).

Here, we explore the hypothesis that OTC sales data can also be used for predicting a disease activity. Towards that end, we developed a model to predict the number of ED flu cases in a region based on TS. We obtain sales information from the National Retail Data Monitor (NRDM) project. NRDM collects daily sales data of 18 OTC categories across the US.[1]

Methods
To create the model, we obtained (1) the number of thermometers sold per day from 1 May to 31 December 2009 in Allegheny County (AC), Pennsylvania (PA), and (2) estimates of the number of daily cases that presented with flu to monitored EDs in AC during the same period. These estimates were produced by our 'Bayesian case detector (BCD)' algorithm, which estimates from an ED report the probability that a given ED patient has flu. We developed a linear regression model with zero time lag to predict flu cases solely from TS; let S and I denote the slope and intercept parameters of the model, respectively.

Presently, the BCD only collects information from a fraction of all EDs in AC ($A = 0.44$), and TS data only cover a fraction of all sales from OTC retail stores in AC ($B = 0.91$). To apply our model to other regions in the US, we used A and B to adjust the regression parameters. Moreover, we also must account for differences in population between AC ($\sim$1.2 million) and the region to which the model is being applied. We obtain the following formula, which assumes that TS and ED visits are proportional to population size (pop) and that our regression model holds:

$$\mathrm{EDFlu(region)} = \mathrm{TS(region)}S(B/A) + (I/A)(\mathrm{pop(region)}/\mathrm{pop(AC)})$$

Given the lack of data with which to validate our model in other regions of PA or the country, we used counts of constitutional symptoms (referred to as constitutional counts, CC) of people who come to EDs as a rough surrogate of ED flu cases. We had such CCs and TS for all counties in PA

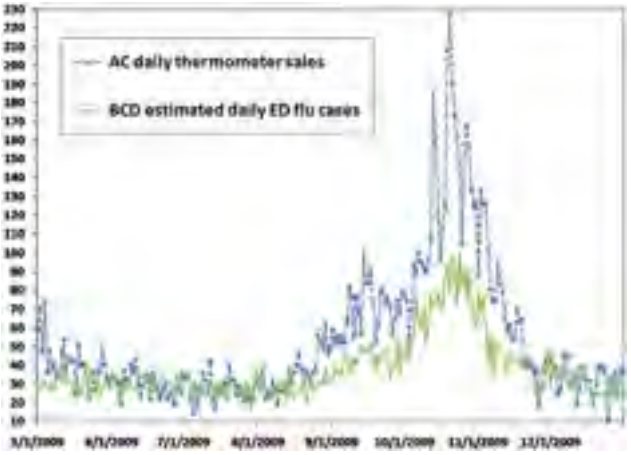

Figure 1 Plots of TS and estimated ED flu cases in AC (1 May–31 December 2009).

in 2009. We trained a linear model to predict CCs for a subset of counties in PA based on those counties' TS in 2009. We applied that model to predict the CCs in another subset of PA counties from their TS. We repeated this experiment 10 times, each time randomly selecting a different subset of counties for training and testing. (Figure 1).

Results

Least squares linear regression applied to TS and counts of ED flu cases (from BCD) produced $S = 0.39$ (95% CI: 0.37, 0.42) and $I = 17.6$ (95% CI: 15.9, 19.2) with R^2 of 0.79 and a cross-correlation (0 time lag) of 0.90. The regressions to predict CCs also had high R^2 values (mean = 0.81) and high correlations (mean = 0.89) between the predictions and actual CCs.

Conclusions

The ability of our regression on TS to predict flu cases (as measured by R^2) shows that a linear relationship between those quantities fits well. Moreover, the regressions done to predict CCs support the validity of predicting outcomes that are related to flu ED visits based on TS. Our results suggest that the use of TS, which are available nationwide daily, has potential to be helpful in estimating ED flu activity.

Acknowledgements

The project was funded by CDC grants P01 HK000086 and 1U38 HK000063-01, NIH grant R01 LM009132, and NSF grant IIS-0911032. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Hogan WR, Wagner MM. Sales of over-the-counter healthcare products. *Handbook of Biosurveillance*, 1st edn. 2006, p 321.
2 Presanis AM, De Angelis D, Hagy A, Reed C, Riley S, Cooper BS, *et al.* The severity of pandemic H1N1 influenza in the U.S., from April to July 2009: a Bayesian analysis. *PLoS Med* 2009;**6:**e1000207.