COACTION PUBLISHING

# In silico surveillance: highly detailed agent-based models for surveillance system evaluation and design

Bryan Lewis[1]*, Stephen Eubank[1], Allyson Abrams[2] and Ken Kleinman[2]

[1]Virginia Tech, Blacksburg, VA, USA; [2]Harvard, Boston, MA, USA

## Objective

To create, implement and test a flexible methodology to generate detailed synthetic surveillance data providing realistic geospatial and temporal clustering of baseline cases.

## Introduction

Modern public health surveillance systems have great potential for improving public health. However, evaluating the performance of surveillance systems is challenging because examples of baseline disease distribution in the population are limited to a few years of data collection. Agent-based simulations of infectious disease transmission in highly detailed synthetic populations can provide unlimited realistic baseline data.

## Methods

Dynamic social networks for the Boston area (4.1 million individuals) were constructed based on data for individuals, locations and activity patterns collected from the real world. We modeled a full season of endemic influenza-like illnesses (ILI), healthcare seeking behavior and a surveillance system for outpatient visits. The resulting in silico surveillance data contain the demographics and complete history of disease progression for all individuals in the population; those who are in a specified surveillance system create a data stream of ILI visits. Outbreaks of influenza are artificially inserted into this surveillance data. Outbreak detection using space-and-time scan statistics was used to analyze the background with and without the inserted outbreaks. The performance of the algorithm was assessed under different levels of coverage and catchment distributions. One hundred unique baseline data sets were generated. Twelve artificial outbreaks were inserted in each. Six different surveillance system designs were assessed.

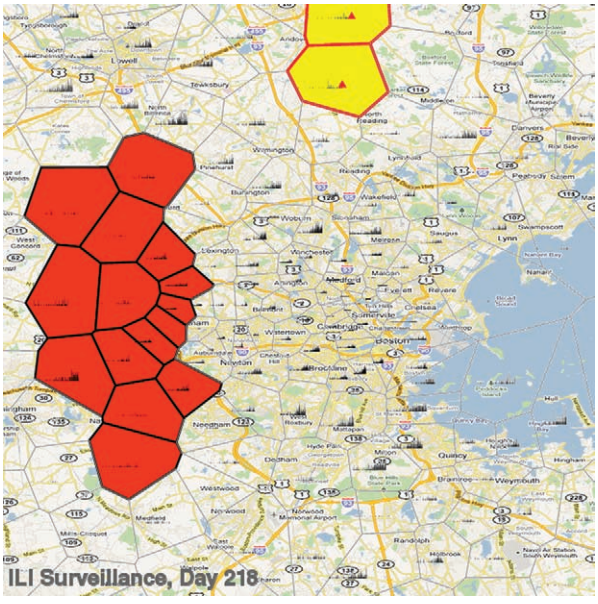


*Fig. 1.* Simulated ILI surveillance data for downtown Boston as captured by simulated surveillance system. Surveillance counts per day centered in each zip code location are shown as histograms within each zip code. Detection of an inserted test outbreak (red triangle) is indicated by red-bordered zip codes and a false-positive outbreak by blackbordered zip codes.

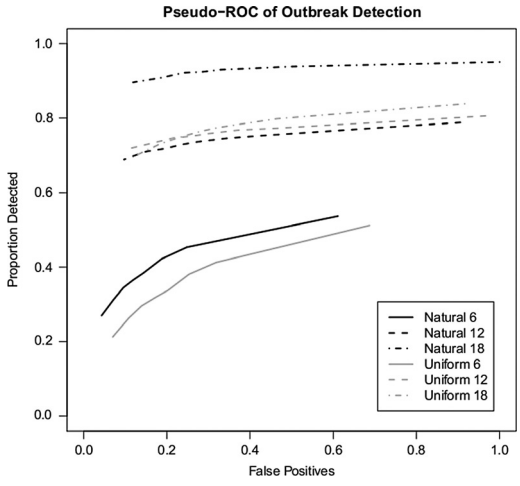


*Fig. 2.* Pseudo-ROC curves of outbreak detection. Proportion detected for each surveillance system vs. proportion of all false-positives identified.

## Results

We present a robust framework for using highly detailed simulations to provide the foundation for evaluating and designing a surveillance system's ability to detect outbreaks. A small demonstration study shows that detection rates varied from 17% to 80% across the different surveillance systems. Increased coverage did not linearly improve detection probability for all surveillance systems. Surveillance systems with uniform coverage of the population did not perform better than one based on a real-world system with nonuniform coverage. Higher coverage improved the timeliness of detection but, for most cases, by only 1 or 2 days on average. Additional results can be found online (http://ndssl.vbi.vt.edu/insilicoSurveillance/).

## Conclusions

Highly detailed simulations of infectious disease transmission can be configured to represent nearly infinite scenarios, making them a powerful tool for evaluating the performance of surveillance systems and the methods used for outbreak detection.

## Keywords

Computer simulation; surveillance evaluation; outbreak detection; SaTScan

*Bryan Lewis
E-mail: blewis@vbi.vt.edu





Citation: Emerging Health Threats Journal 2011, **4**: 11027 - DOI: 10.3402/ehtj.v4i0.11027