The Effect of Inferring Work Location from Home Location in Performing Bayesian Biosurveillance

Chris Garman, M.Sc., Weng-Keen Wong, Ph.D., Gregory F. Cooper, M.D., Ph.D.

RODS Laboratory, Center for Biomedical Informatics, University of Pittsburgh

OBJECTIVE

We describe and evaluate an extension to spatial biosurveillance algorithms that allows a patient's work zip code, which is often missing, to be inferred from the patient's home zip code.

BACKGROUND

Incorporating spatial information can improve the performance of outbreak detection algorithms (1). Spatial detection algorithms typically use the patient's home zip code as an approximation to the point of exposure. Of course, a patient may not necessarily be at home during a bioterrorist attack. If an attack occurs at work, the work zip code is a much better approximation. However, unlike the home zip code, which is usually available in ED data, the work zip code of a patient is often missing. In (2), the authors deal with this problem by first locating significant clusters of increased morbidity in time and space using the patients' home addresses and then finding the center of minimum distance to these clusters.

METHODS

Another approach to dealing with missing work zip codes is to estimate the probability that a person living in zip code *X* works in zip code *Y*. This probability distribution can be learned from historical data or obtained from census information. With this distribution, we can infer a patient's work zip code given her home zip code. This strategy applies for any spatial outbreak detection algorithm that uses a probabilistic approach to deal with uncertainty. We extended the PANDA algorithm (3) to have the ability to infer a patient's work zip code given her home zip. The new model examines ED data for a specific spatial pattern in the home locations of the patients and alternatively in the inferred work locations.

We evaluate our approach using a simulator (4) to generate 30 simulated large scale anthrax releases that occur at the work locations of the patients. In the *work zip* experiment, the PANDA algorithm is (idealistically) given access to the work zip codes of all the patients. In the *home zip* experiment, the PANDA algorithm, using a model without the ability to infer the work zip code, is only given access to the home zip codes of the patients. Finally, in the *home zip to work zip* experiment, the PANDA algorithm (using a model with the ability to infer the probability of work zip codes from home zip codes) is only given access to the patients' home zip codes.

RESULTS

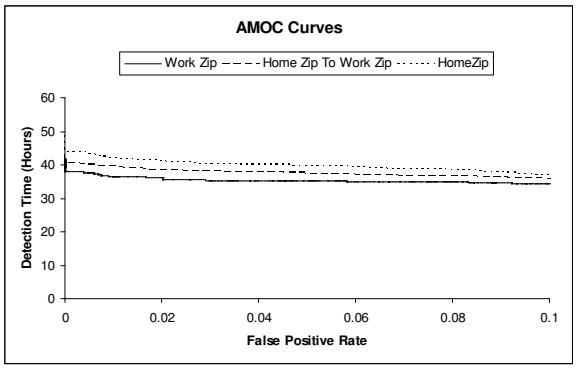

Figure 1: The AMOC curves of the three experiments, averaged over 30 simulated attacks. At a zero false positive rate, the detection times are 37.8, 44.0, and 40.9 hours after the anthrax release for the *work zip*, *home zip*, *and home zip to work zip* experiments respectively.

CONCLUSIONS

The results indicate that if a large scale anthrax attack happens at work, inferring patients' work zip codes from their home zip codes improves the performance of the detection algorithm on the simulated attacks over just using the home zip codes of the patients as the point of infection.

REFERENCES

(1) Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. Biomed Inform 2005; 38 (2):99-113.

(2) Green MS, Kaufman Z. Syndromic surveillance for early location of bioterrorist incidents outside of residential areas. In Proceedings of the National Syndromic Surveillance Conference [CD-ROM]. Boston, MA: Fleetwood Multimedia, Inc.; 2004.

(3) Cooper GF, Dash DH, Levander JD, Wong WK, Hogan WR, Wagner MM. Bayesian Biosurveillance of Disease Outbreaks. In Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence. Banff, Canada: AUAI Press; 2004. 94-103 p.

(4) Hogan WR, Cooper GF, Wallstrom GL, Wagner MM. The Bayesian aerosol release detector. In Proceedings of the National Syndromic Surveillance Conference [CD-ROM]. Boston, MA: Fleetwood Multimedia, Inc.; 2004.