Emerging Health Threats Journal 2011, **4**:s22. **doi:** 10.3134/ehtj.10.022

www.eht-journal.org

ABSTRACT

Evaluating the performance of two alternative geographic surveillance schemes

G Fairchild, A Segre, P Polgreen, and G Rushton

Computational Epidemiology Research Group, Department of Computer Science, University of Iowa, Iowa City, IA, USA
E-mail: gcfairch@gmail.com

Objective

This paper describes an experiment to evaluate the performance of several alternative surveillance site placement algorithms with respect to the standard influenza-like illness (ILI) surveillance system in Iowa.

Introduction

ILI data is collected by an Influenza Sentinel Provider Surveillance Network at the state (Iowa, USA) level. Historically, the Iowa Department of Public Health (IDPH) has maintained 19 different influenza sentinel surveillance sites. Because participation is voluntary, locations of the sentinel providers may not reflect optimal geographic placement. This study analyzes two different geographic placement algorithms—a maximal coverage model (MCM) and a K-median model.[1] The MCM operates as follows: given a specified radius of coverage for each of the n candidate surveillance sites, we greedily choose the m sites that result in the highest population coverage. In previous work, we showed that the MCM can be used for site placement.[2] In this paper, we introduce an alternative to the MCM—the K-median model. The K-median model, often called the P-median model in geographic literature, operates by greedily choosing the m sites which minimize the sum of the distances from each person in a population to that person's nearest site. In other words, it minimizes the average travel distance for a population.

Methods

This project is split into two phases—a calculation phase and a validation phase. In the first phase, we developed a user-friendly web-based calculator to help public health entities locate sites on their own. In the second phase, we compared and analyzed the placements of sites to show that the two models implemented in the web calculator choose 'better' sites *de novo* for Iowa than the existing 19 sites hand-picked by the IDPH. First, we generated 19 sites *de novo* using each algorithm. Then, we used 8 years of statewide Medicaid billing data to simulate the spread of influenza across the state of Iowa. In total, there were approximately 2 million cases with ILI-related ICD9 codes present in this Medicaid database. We used two different probability functions, an exponentially decreasing function relating to patient distance from site, and the Huff probability model, a model commonly used in geography, to probabilistically measure the likelihood of a case being 'noticed' by a surveillance site (and thus having their visit reported back to the IDPH). The Huff model operates by computing the probability based both on the distance to a site and the 'attractiveness' of a site, which, for this experiment, is the population coverage of that site.

Results

	Existing	*MCM*	*K-median*
Exponential	16.0%	12.0%	19.3%
Huff	32.3%	37.4%	41.3%

The Table above shows the percentage of Medicaid cases detected during the simulation for each model and for each probability function. The 19 sites calculated *de novo* using the K-median model capture 3.3% more cases than the 19 existing sites when using the exponentially decreasing probability function based on distance. Additionally, using the commonly used Huff probability model, the same sites achieve 9% better detection than the existing sites. Although the sites calculated using the MCM detect fewer cases than the existing sites using the distance-based probability function, they capture 5.1% more cases using the Huff model.

Conclusion

We have described a systematic, intuitive and easy to use method by which surveillance sites may be located in a given geographic region. We offer two models for site placement. One model, the maximal coverage model, greedily chooses sites which maximize total population coverage. The other model, the K-median model, greedily chooses sites which minimize the average travel distance for a given population. When simulating 8 years of ILI Medicaid cases in Iowa against the chosen sites, we found that we are able to generate sites that outperform the existing sites in Iowa in terms of cases reported when compared using two different probabilistic models for patient visits.

To generate the best results from a disease surveillance network, site placement is of the utmost importance, and the methods we describe help to ensure that sites are located as well as possible.

Acknowledgements

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 Church RL. Location modeling and GIS. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds). *Geographical Information Systems, Principles, Techniques, Applications and Management*. 2nd edn. New York: Wiley, 1999, pp 293–303.
2 Polgreen P, Chen Z, Segre AM, Harris ML, Pentella MA, Rushton G. Optimizing influenza sentinel surveillance at the state level. *Am J Epidemiol* 2009;**170:**1300–6.