Benchmarking Temporal Surveillance Techniques

James Dunyak[1], Ph.D., Mojdeh Mohtashemi[1,2], Ph.D., Martin Kulldorff[3], Ph.D.

[1]*The MITRE Corporation;* [2]*MIT AI and CS Lab;* [3]*Harvard Medical School and Harvard Pilgrim Health Care*

OBJECTIVE

This paper compares the robustness and performance of three temporal surveillance techniques using a twofold approach: 1) a unifying statistical analysis to establish their common features and differences, and 2) a benchmarking on respiratory, influenza-like illnesses, upper GI, and lower GI complaint time series from the Harvard Pilgrim Health Care (HPHC).

BACKGROUND

Benchmarking of temporal surveillance techniques is a critical step in the development of an effective syndromic surveillance system. Unfortunately, holding "bakeoffs" to blindly compare approaches is a difficult and often fruitless enterprise, in part due to the parameters left to the final user for tuning. In this paper, we demonstrate how common analytical development and analysis may be coupled with realistic data sets to provide insight and robustness when selecting a surveillance technique.

METHODS

The algorithm development and analysis begins with a locally stationary Poisson arrival model for the patients. An outbreak is characterized by an increase in the daily Poisson mean, resulting in stochastic outbreaks. The vast majority of the syndromic surveillance literature addresses the detection of step increases in patient levels. We address the more general issue of detecting monotonic increases in mean as characteristic of the early stage of an outbreak following a classic susceptible-exposed-infected-removed model.

This time-varying mean estimate is applied within a detection window of length T, nominally of 3-14 days for early detection. This moving detection window scans the time series in real-time, as measurements are available, to detect a current outbreak.

A likelihood ratio provides a common basis for three different algorithms, which we compare in this paper. 1) A uniformly most powerful (UMP) test can be developed as a matched filter for an entire class of outbreak profiles. This viewpoint, after application of the central limit theorem, is applied in [1] and [2]. 2) Applying a generalized likelihood ratio test (GLRT) to optimize window length results to detect step increases in mean results in the G-surveillance statistic [3], which is itself an adaptation of the widely used scan statistic [4]. 3) Finally, a new approach uses the GLRT formulation in conjunction with a monotonic regression viewpoint for the time varying mean.

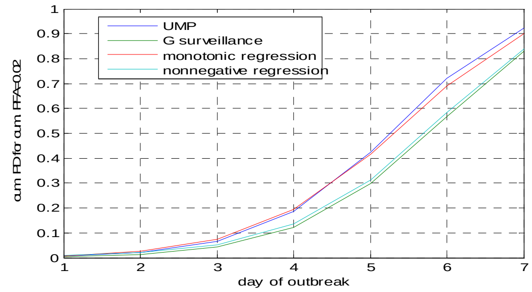

Fig. 1 – Cumulative probability of detection at each day of an exponentially growing outbreak, with cumulative false positive at day seven=0.02.

RESULTS

All three viewpoints are compared theoretically and then benchmarked using five years of daily visits in the Boston area of the HPHC data. In particular, the tradeoff between timeliness, robustness, and maximum detection power is established in a wide number of settings. The numerical results, from both simulation and application of the HPHC experimental time series, are explained theoretically in terms of signal-to-noise-ratio. Figure 1 illustrates some of the results. The benefit of the monotonic regression viewpoint is clearly established when the trajectory of the outbreak is not known *a priori*.

CONCLUSIONS

Comparing and selecting temporal surveillance techniques is a complicated task due to the wide variety of techniques in the literature. In this paper, we demonstrate an approach for this process based on a common analytical development and benchmarking with common data sets.

REFERENCES

[1] Reis BY, Pagano M, Mandl KD, Using temporal context to improve surveillance, PNAS. 2003;100(4):1961-1965.

[2] Dunyak, J., Mandl, K., and Mohtashemi, M. A Binomial Model of Transients in Daily Emergency Department Visits for Detecting Infectious Disease Outbreaks. Submitted to IEEE Trans. On Biomedical Engineering. Under review.

[3] Wallenstein, S. and Naus, J. Scan Statistic for Temporal Surveillance for Biological Terrorism. Morbidity and Mortality Weekly Report. 2004; v. 53, p. 74-78.

[4] Naus J. The distribution of the size of the maximum cluster of points on a line. J Am Stat Assoc 1965;60:532--38.