# Minimizing False Alarms in Syndromic Surveillance

**William Peter, Ph.D., Amir H. Najmi, D.Phil., Howard Burkom, Ph.D.**
*The Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723*

## OBJECTIVE

This paper describes a method of avoiding false alerts in automated syndromic surveillance algorithms which monitor the temporal relationship between a particular monitored syndrome (the "target") in relationship to other reference healthcare data streams (the "context").

## BACKGROUND

The identification of public health threats or disease outbreaks relies on opportunistic data streams that vary greatly across geographic regions, time periods, and other factors [1]. Most surveillance systems are vulnerable to dramatic and unpredictable shifts in the healthcare data they monitor [2] since anomalous counts in syndromic data streams can be biased by changes in data collection methods, seasonality, etc. For example, a sudden jump in influenza-like illness diagnoses in a population might be caused by an increase in the monitored population size.

One method of avoiding false alerts in syndromic monitoring is to use suitably-normalized syndromic counts (i.e., ratios or proportions) instead of the individual counts themselves. The use of such normalized time series lowers the number of false alerts since it only compares the changes in the target time series (the "numerator") to corresponding changes in the background (the context data stream, or "denominator").

In many cases, the choice of a suitable "denominator", or reference stream, is not obvious. One recent approach [2] has been to use all possible pair-wise ratios in an epidemiological network of $N$ healthcare data streams. The $N^2$ scaling of the size of this network, however, is quite formidable.

## METHODS

Our approach to solve the "denominator problem" is to use information-theoretic techniques to determine under what conditions proportions are preferable to counts, and what an appropriate context time series for a given target syndrome might be. Given two data streams $\{x(t)\}$ and $\{y(t)\}$, their mutual information $I(X;Y)$ provides a general measure for their mutual dependency, or nonlinear correlation. If two data streams have a high mutual information, the use of proportions to clarify an outbreak signal is justified. In the opposite case the use of proportions will not clarify alert identification.

To this end, assorted pairs of syndromic data series with variable mutual information have been compared. The simulations involve introducing lognormal spikes in the daily counts of the numerator series to simulate an alert. This alert is then identified by the corresponding decrease in mutual information between the two series and also by its respective peak in the plot of proportional values of the two series. ROC curves will also be presented.

## RESULTS

One advantage of our method is the expected clear-cut identification of meaningful alerts based on requiring the target and context time series to have mutual information above a given threshold. The approach here complements the proportional model networks of Reis, Kohane, and Mandl [2], and may be used to decide which target/context ratios will add value to a multi-source network, thus reducing the size and complexity of the calculations.

A general and potentially more robust detection algorithm based on the instantaneous loss in mutual information will also be presented. If $\Delta I$ is the change in mutual information between a target syndrome series and its context, anomalous changes in the "speed of mutual information loss", given quantitatively by $-\Delta I/\Delta t$, seems to be a more general and robust alert detector than thresholding the proportions themselves.

## CONCLUSIONS

Appropriately conditioning and combining syndromic data streams is important to achieve the signal-to-noise gains necessary for increased sensitivity in biosurveillance. Information-theoretic methods, little used in disease surveillance up to now, will be applied to gain practical guidance toward this goal.

## REFERENCES

[1] Burkom H, Alerting Algorithms for Biosurveillance, in *Disease Surveillance* (Eds: Lombardo J S and Buckeridge D L), Wiley-Interscience, Hoboken, NJ 2007.

[2] Reis B Y, Kohane I S., and Mandl K D. An Epidemiological Network Model for Disease Outbreak Detection, PLOS Medicine (to appear, 2007).

Further Information:
Bill Peter, bill.peter@jhuapl.edu