

# Preparing Biosurveillance Data for Classic Monitoring

Thomas Lotze<sup>1</sup>, Sean P. Murphy<sup>2</sup> and Galit Shmueli<sup>1,3</sup>

<sup>1</sup>Applied Math and Scientific Computation Program, Univ of Maryland College Park

<sup>2</sup>The Johns Hopkins University Applied Physics Laboratory

<sup>3</sup>Decisions & Information Technologies Dept, Smith School of Business, Univ of Maryland

## Objective

Statistical process control (SPC) charts are widely used in disease surveillance. The charts are very effective when monitored data meet the requirements of temporal independence, stationarity, and normality. However, when these assumptions are violated, the SPC charts will either fail to detect special cause variations or will alert frequently even in the absence of anomalies. Currently collected biosurveillance data contain predictable factors such as day-of-week effects, seasonal effects, holidays, autocorrelation, and global trends that cause the data to violate these assumptions. This work (1) describes a set of tools for identifying such explainable patterns and (2) examines several data preconditioning methods that account for these factors, yielding data better suited for monitoring by traditional SPC charts.

## Background

Modern surveillance systems use SPC charts such as Cumulative Sum (CuSum) and Exponentially Weighted Moving Average (EWMA) charts for monitoring daily counts of such quantities as ICD-9 codes from ED visits, sales of medications, and doctors' office visits. The working assumption is that such pre-clinical data contain an early signature of disease outbreaks, manifested as an increase in the count levels. However, the direct application of SPC charts to the raw counts leads to unreliable performance. A popular statistical solution is to precondition the data before applying the charts by modeling or removing explainable patterns from the data and then monitoring the residuals. Although the general idea is common practice, the specifics of how to identify the existing explainable components and how to account for them are domain-specific. Therefore, we seek to present a set of modeling and data-driven tools that are useful for syndromic data.

## Methods

The first part of this study evaluated numerous techniques for identifying explainable patterns including graphical methods (e.g., zoomed time plots, autocorrelograms, probably plots) and summary statistics (e.g., skewness measures, summaries by day-of-week). The second part examined four preconditioning methods: linear regression models, ratio to moving average indexes, differencing, and Holt-Winters exponential smoothing. Authentic syndromic data for this study came from two sources: 2 years of daily sales of 7 categories of medications

from a grocery chain in the Pittsburgh area [1] and 3 years of 35 series of daily patient arrival counts at emergency departments in an unspecified metropolitan region. The tools from part 1 were used to uncover explainable patterns in the data, and also to evaluate the methods of part 2.

## Results

Figure 1 compares the application of a CuSum chart to raw counts of cough medication sales (top) vs. to the preconditioned series using the different methods. As the red dots indicate alarms, the preconditioning is clearly a necessary step. Similar results and additional evaluations were performed for all other series. The day-of-week effect is a major cause of control chart alerts and, in some series, so is seasonality.

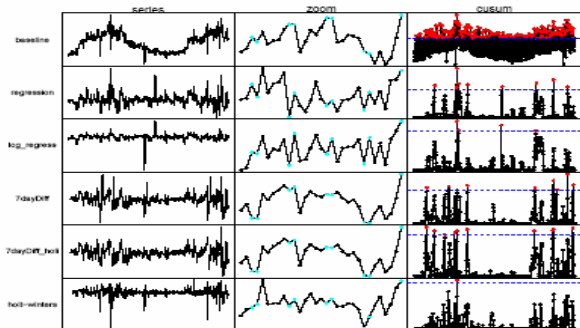


Figure 1: Raw (top) and five preconditioned versions of daily cough medication sales series, with zoomed-in view (center) and CuSum chart applied (right)

## Conclusions

More than recommending one particular method, we believe that some preconditioning technique must be used to remove weekly and yearly seasonality from syndromic data to help it approach normality in order to achieve better results using control charts. Unless this is done, the seasonal patterns will dominate both the data and the alerts. Further work remains to investigate how each preconditioning method impacts different outbreak signals and the outbreak detection performance of traditional SPC monitoring methods.

## References

- [1] Goldenberg, A., Shmueli, G., Caruana, R. A., and Fienberg, S. E. (2002), "Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-Counter Medication Sales", *Proceeding of the National Academy of Sciences*, vol. 99, Issue 8, pp. 5237-5240.