Data-Adaptive Multivariate Control Charts for Routine Health Monitoring

Sean P. Murphy, Howard S. Burkom, Ph.D., Galit Shmueli, Ph.D.

The Johns Hopkins University Applied Physics Laboratory and the Decision and Information Technologies Department, Robert H. Smith School of Business, U. of Maryland College Park

OBJECTIVE

This paper investigates the use of data-adaptive multivariate statistical process control (MSPC) charts for outbreak detection using real-world syndromic data. The widely used EARS [1] methods and other adaptive implementations assume implicitly that nonstationarity and/or the lack of historic data preclude the conventional Phase I/Phase II approach of SPC. This work examines that assumption formally by evaluating and comparing the false alarm rates and sensitivity of adaptive and non-adaptive MSPC charts applied to simulated outbreaks injected into both deseasonalized and raw data.

BACKGROUND

Classical process control charts have been used effectively for industrial quality control for several decades and have many potential applications in hospital surveillance [2]. In this field, the processes examined are assumed to be normal with a constant mean and variance and thus amenable to a 2-stage approach: first, baseline data are analyzed to derive statistical properties expressed as chart parameters, and then the chart is prospectively applied. In sharp contrast, the myriad of processes responsible for creating syndromic data—human behavior, immune physiology, the health care system, to name just a few—often exhibit explicit trends, cyclical behaviors, and statistical properties that change over time. As researchers begin to apply multivariate charts to health surveillance [3], both preconditioning of the data and adaptation of the algorithms become essential.

METHODS

This investigation, using data preconditioning, algorithm adaptation, and performance analysis, examined authentic syndromic time series derived from three data sources from each of five US cities using both respiratory and gastrointestinal data counts. For data preconditioning, we applied several local deseasonalization techniques—exponential smoothing, subtraction of the count from seven days ago [4] and additive and multiplicative versions of the classic "ratio-to-moving-average" method—to remove seasonal effects that degrade chart performance. To adapt the algorithms to a changing mean and covariance, we implemented each with outlier removal and a moving four week baseline with a 2-day guardband for estimating chart parameters. The adapted MSPC charts included multivariate versions of a cumulative sum, exponentially-weighted moving average (M-EWMA) and Hotelling's T^2 charts. Performance analysis occurred in two phases. First, data intervals containing outbreaks, identified by an independent group of medical epidemiologists, were removed from the data and a comparative measure of false alert rates was determined by considering any anomalies in the remaining data to be spurious (Fig 1). Second, two types of stochastic signals were injected into the syndromic data and each method's sensitivity was computed for detecting these simulated outbreaks as a function of the background alert rate.

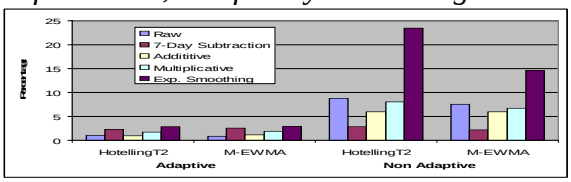

Figure 1—Average percentage of non outbreak days that false alarmed for a set threshold using both adaptive and non adaptive MSPC charts using various deseasonalization techniques.

RESULTS

Analysis of the original syndromic data found a pronounced "day of week" effect that deseasonalization techniques could remove as evidenced by qualitative examination of the post-processed power spectra. The data's seasonal variations had a significant impact on both the performance of the MSPC charts as measured by their sensitivity to simulated outbreak detection and their robustness as measured by false alarm rates during non-outbreak days. This performance degradation could be mitigated by adapting the syndromic data, the traditional MSPC charts, or both.

CONCLUSIONS

Although the specific determination of how and whether to adapt quality control techniques should be made based on the statistical properties of the data, we found that both data deseasonalization and chart adaptation enhanced MSPC chart performance when applied to syndromic surveillance.

REFERENCES

[1] Hutwagner LC. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urb Health 2003;80:i89–96.

[2] Benneyan, JC, "Statistical Quality Control Methods in Infection Control and Epidemiology.: Intro and Basic Theory", invited series, Stat Hosp Epi, Inf Cont & Hosp Epi,19(3):194-214, 1998.

[3] Rogerson PA and I. Yamada. "Monitoring Change in Spatial Patterns of Disease: Comparing Univariate and Multivariate Cumulative Sum Approaches." Stat Med, 23 (14), 2004, 2195-2214.

[4] Muscatello D, "An adjusted cumulative sum for count data with day-of-week effects: application to influenza-like illness." 2004 Syn Surv Conf, Boston MA, Nov 3-4, 2004.

Further Information: Sean Murphy, sean.murphy@jhuapl.edu