for detection of syndromes in individuals using a heuristic approach, normalization using Kalman Filtering, and epidemic detection using a Gaussian Bayesian belief network. Based on our experience with these prototypes, we are beginning to formally model surveillance knowledge requirements, implement additional PSMs, and develop an evaluation framework.

SECTION IV: STATISTICAL METHODOLOGIES

A Cumulative Sum Approach to Syndromic Surveillance in Geographic Regions

Peter A. Rogerson
Department of Geography, University at Buffalo, Buffalo, New York

A primary objective of syndromic surveillance is to find as quickly as possible any significant increase in the rates of syndromes of interest. One way to achieve this is to use cumulative sum methods; such methods are commonly used in industry to detect unwanted changes in industrial processes, and they fall within the more general category of statistical process control methods. Cumulative sum (CUSUM) methods operate by accumulating deviations between observations and expectations. When these cumulated deviations exceed some predefined threshold, an alarm is sounded to indicate an increase in the mean of the underlying variable of interest. Daily data were available on the number of visits made by patients residing within given census tracts to clinicians in the Boston, Massachusetts, area for lower respiratory infections. Expected counts were modeled for the period 1996–1998 using logistic regression and using month, weekday/weekend, and a time trend as covariates. In general, weekends and summer months result in lower odds of office visits in comparison with weekdays and winter months. The observed counts were then compared with the expected counts using the cumulative sum method. The small daily counts suggest that a Poisson CUSUM be employed, for which the expectations vary over time. The method outlined can be extended to account for the fact that observed counts typically exhibit more variability than a Poisson model would suggest. The current approach detects increases in specific regions; a next step is to allow for the possibility that small clusters of geographic regions witness increases in rates.

An Elliptic Spatial Scan Statistic and Its Application to Breast Cancer Mortality Data in Northeastern United States

Martin Kulldorff, Lan Huang, and Linda Pickle
University of Connecticut, National Cancer Institute

The spatial scan statistic is commonly used for geographical cluster detection, cluster evaluation, and disease surveillance. Recent use includes daily analyses of syndromic emergency room data for the early detection of disease outbreaks in New York City. Whether an outbreak is due to a natural cause or a bioterrorism attack, this system enables city health officials to investigate the outbreak as early as possible and, if necessary, to rapidly implement disease prevention and control measures. When applying the spatial scan statistic, it is natural to use a circular scanning window to define the potential cluster areas since the circle is the most compact shape that can be obtained. Other shapes are also possible, such as

ellipses. An elliptic scanning window could provide higher power if the true cluster shape is noncircular, which one would often expect to be the case. Here, we describe and illustrate the use of an elliptic spatial scan statistic and apply it to breast cancer mortality data in the northeastern United States. The selection of the elliptic shapes and angles are discussed and comparisons are made between circular spatial scan statistic and elliptic scan statistic by implementing power study.

Geographic and Network Surveillance for Arbitrarily Shaped Hotspots—Next Generation of Potential Outbreak Detection and Prioritization System

G. P. Patil, W. L. Myers, C. Taillie, and D. Wardrop
Pennsylvania State University

We present a version of the spatial scan statistic that is intended to address the following shortcomings of circle-based scans:

- Circles are able to capture only compactly shaped clusters. In many applications, clusters can have a very irregular shape. Cylindrical zones can yield poor hot spot delineation in a space-time scan.
- The circle-based zonation relies on Euclidean distance and is inappropriate for data defined along a network.
- The spatial scan statistic yields a maximum likelihood point estimate for the hot spot, but provides no assessment of the uncertainty or variability. One would like to have alternative plausible delineations of the hot spot expressed as a hot spot confidence set.

Our version of the scan statistic employs the upper level set (ULS) of the response rate defined over the cells of a tessellation (or over the nodes of a network). Attractive features of the ULS scan statistic include

- Identification of arbitrarily shaped clusters
- Data-adaptive zonation of candidate hot spots
- Applicability to data on a network
- Provision of both a point estimate and a confidence set for the hot spot
- Use of hot spot membership rating to map hot spot boundary uncertainty
- Computational efficiency
- Applicability to both discrete and continuous responses
- Identification of arbitrarily shaped clusters in the spatial-temporal domain

We also present a prioritization innovation. It lies in the ability for prioritization and ranking of hot spots based on multiple indicator and stakeholder criteria, using partial order sets, without having to integrate indicators into an index.

Biosurveillance Applying Scan Statistics With Multiple, Disparate Data Sources

Howard S. Burkom[1] and Eugene Elbert[2]
[1]*National Security Technology Department, Johns Hopkins Applied Physics Laboratory,*
[2]*Walter Reed Army Institute of Research*

Researchers working on the Department of Defense Global Emerging Infections System (DoD-GEIS) pilot system, the Electronic Surveillance System for the Early Notification of

Community-Based Epidemics (ESSENCE), have applied scan statistics for early outbreak detection using both traditional and nontraditional data sources. These sources include medical data indexed by *International Classification of Diseases, 9th Revision* (*ICD-9*) diagnosis codes as well as less-specific, but potentially timelier, indicators such as over-the-counter remedy sale totals and school absenteeism records. Early efforts have employed the Kulldorff scan statistic as implemented in the SaTScan software of the National Cancer Institute. A key obstacle is that the input data streams are typically based on time-varying factors such as consumer behavior rather than simply on the populations of the component subregions. Both modeling and recent data have been used to obtain background spatial distributions. Data analyses have provided guidance for determining baseline periods to avoid excessive clustering. We used a simple covariate approach to combining data sources and are evaluating alternative fusion methods in a test bed setting. Experience with this methodology has included combinations of data sources for both retrospective studies of known outbreaks and surveillance of high-profile events of concern to local public health authorities. We have developed capability to test the detection performance of scan statistics as an outbreak unfolds. Spatial and temporal epicurve simulations are used to inject cases into the various streams of authentic data to enable day-by-day performance analysis.

Space-Time Disease Map Surveillance With Extensions to Bioterrorism

Andrew B. Lawson

Department of Epidemiology and Biostatistics, Norman J. Arnold School of Public Health, University of South Carolina

There is increasing interest in health surveillance, particularly syndromic surveillance, or early detection of adverse health effects. This interest now focuses on the need to monitor in real time or near real time conjointly such indicators as pharmaceutical sales, job absences, and physician or emergency room attendances to assess whether there is any incipient evidence of an outbreak or other sustained adverse health event. This interest stems from the perceived threat of biochemical and virus-based attacks following the terrorist attacks of September 11, 2001. A number of concerns arise when the statistical issues surrounding space-time map surveillance are considered. In this poster, I describe some of the important issues. The important ingredients of good syndromic statistical algorithms are as follows:

- The model used for the map should be general and not overparameterized as changes over time in the map are to be detected.
- The model should be sensitive to allow early detection of important changes.
- The model should allow multivariate and multifeature surveillance and be adaptable when increased knowledge is available.

Maps should be examined for changes in trends, clustering, and discontinuities, and any good system should be able to find these quickly within a map. Currently, many methods have been introduced to surveillance systems that were not designed for the surveillance task. The design of optimal surveillance methods will be a future need. Currently, we can apply standard Bayesian hierarchical models to evolving count data sets or can conceive of a modeling framework for which data are only available by searching in real time (e.g., mobile recording units). There is a need to greater development and evaluation of statistical algorithms for the surveillance task.

Bibliography Lawson A. Space and space-time disease map surveillance. In: Brookmeyer R, Stroup D, eds. *Monitoring the Health of Populations: Statistical Methods for Public Health Surveillance*. Oxford, UK: Oxford University Press; 2003.

Large scale: surveillance. In: Lawson AB. *Statistical Methods in Spatial Epidemiology*. New York, NY: Wiley; 2001.

Syndromic Surveillance and Risk Management Using Multiitem Gamma Poisson Shrinker

Ana Szarfman
QMR/OB/OPaSS, CDER, Food and Drug Administration

Scientists at the Food and Drug Administration (FDA) have been exploring DuMouchels' automated and rapid Bayesian data mining techniques to systematically and rapidly screen the huge MedWatch database of voluntary reports for possible events of concern with existing and new drugs. The Multiitem Gamma Poisson Shrinker (MGPS) algorithm computes signal scores (SS) (adjusted ratios of observed-to-expected drug event reports, O/E) from application of a statistical model with stratification to control for potential confounding by age, gender, or reporting period. MGPS systematically generates reliable, consistent, redundant, and replicable early signals from the data while minimizing random patterns. Signals are generated without linkage to the corresponding, unfeasible to match systematically, external exposure data or adverse event background information. MGPS computes SS for pairs and for higher order (e.g., triplet, quadruplet) combinations of drugs and events that are significantly more frequent than their pairwise associations would predict. Only a relatively small proportion (3.4%) of all distinct drug-event pairs in the database result in large SS (>2.0). These signals capture a high proportion (23%) of the total number of drug-event pairs reported, greatly facilitating more focused follow-up, evaluation, and risk management. In addition to detecting possible serious single-drug adverse event problems, MGPS detects possible synergistic interactions between drugs (drug interactions) and among adverse events (syndromes). Examples of syndromes automatically detected without the need of case definitions include the following death-associated quadruples having rhabdomyolysis as an event: blood creatine phosphokinase increased–blood myoglobin increased–myoglobinuria present–rhabdomyolysis, with 5 reports, mostly explained by pairwise interactions; and hypotension–myocardial infarction–renal failure acute–rhabdomyolysis, with 4 reports, mostly unexplained by pairwise interactions.

A Comparison of Military Surveillance Systems for Early Detection of Naturally Occurring and Bioterrorism-Based Epidemics of Febrile Respiratory Illness

Nicola Marsden-Haug,[1] Eugene Elbert,[1] Anthony Hawksworth,[2] Virginia Foster,[1] and Julie Pavlin[1]
[1]*Department of Defense–Global Emerging Infections System, Walter Reed Army Institute of Research,* [2]*Naval Health Research Center, San Diego, California*

To evaluate the performance of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE), an electronic health indicator surveillance system, it was compared to a more traditional, active system operated by the Naval Health Research Center (NHRC). The NHRC system has research assistants at military treatment facilities (MTFs) to record the number of recruits who meet a case definition of febrile respiratory illness (FRI). Data are forwarded to the NHRC, where weekly rates of FRI per MTF are calculated. The ESSENCE system groups already collected diagnostic data in the form of *International Classification of Diseases, 9th Revision* (*ICD-9*) codes into syndromic categories. The ESSENCE server receives updated *ICD-9* data every 8 hours; however, submission of this information from the MTFs ranges from 1 to 3 days. A subset of codes for diagnoses and symptoms that could constitute an FRI was generated from the ESSENCE data. The correlation of weekly FRI case detection between the two systems from June 1998 to January 2002 was evaluated in nine MTFs. The surveillance data from the four MTFs with the highest degree of correlation was further compared to assess whether both systems

detected the same epidemics and how quickly they did so. When using autoregessive modeling, ESSENCE did not always detect high peaks following slow increases. Other statistical methods are being developed to correct this omission. Despite this, there is strong evidence from some MTFs that ESSENCE can capture surveillance trends similar to those seen in the distinctly different type of system operated by NHRC.