

# Syndromic Surveillance

An Article for

*The Encyclopedia for Quantitative Risk Assessment*

Ronald D. Fricker, Jr.  
Naval Postgraduate School  
Monterey, California, USA

December 25, 2006

**Abstract:** Syndromic surveillance is the regular collection, analysis, and interpretation of real-time and near-real-time indicators of diseases and other outbreaks by public health organizations. Motivated by the threat of bioterrorism, syndromic surveillance systems are being developed and implemented around the world. In a 2004 systematic review of publicly available information, 115 surveillance systems were identified, of which 29 were found that were designed specifically for detecting bioterrorism. In spite of their development, it is unknown how effective these systems will be at quickly detecting a bioterrorism attack. However, under the rubric of electronic biosurveillance, the goal of some of these systems has recently been expanded to include both early event detection and situational awareness, so that the focus is not simply on detection, but also on response and consequence management. Regardless of their utility for detecting bioterrorism, there seems to be consensus that these biosurveillance systems are likely to be useful for detecting and responding to natural disease outbreaks such as seasonal and pandemic flu, and thus they have the potential to significantly advance and modernize the practice of public health surveillance.

**Keywords:** biosurveillance, bioterrorism, public health, early event detection, situational awareness

*Syndromic surveillance* has been defined as “the ongoing, systematic collection, analysis, interpretation, and application of real-time (or near-real-time) indicators of diseases and outbreaks that allow for their detection before public health authorities would otherwise note them.” [1] It has also been defined as “...surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response.” [2] [3] These definitions focus on a number of concepts important to syndromic surveillance:

1. Syndromic surveillance is public health surveillance, not military, regulatory, or intelligence surveillance.
2. The surveillance is based on the ongoing, routine collection of “health-related” data, such as counts of individuals coming into

medical facilities, over-the-counter medication sales, and aggregate laboratory test results, that is generally collected for some purpose other than surveillance.

3. The data and associated surveillance are intended to provide leading indicators that “precede diagnosis” or ‘case’ confirmation with the goal of providing early warning of an outbreak.
4. The system must provide timely signals that trigger “further public health response” while simultaneously minimizing the number of false alarms. [3]

A *syndrome* is “A set of symptoms or conditions that occur together and suggest the presence of a certain disease or an increased chance of developing the disease.” [4] In the context of syndromic surveillance, a syndrome is a set of non-specific pre-diagnosis medical and other information that may indicate the release of a bioterrorism agent or natural disease outbreak. (See, for example, Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents. [5]) The data in syndromic surveillance systems may be clinically well-defined and linked to specific types of outbreaks, such as groupings of ICD-9 codes from emergency room “chief complaint” data, or only vaguely defined and perhaps only weakly linked to specific types of outbreaks, such as over-the-counter sales of cough and cold medication or absenteeism rates.

Since its inception, syndromic surveillance has mainly focused on *early event detection*: gathering and analyzing data in advance of diagnostic case confirmation to give early warning of a possible outbreak. Such early event detection is not supposed to provide a definitive determination that an outbreak is occurring. Rather, it is supposed to signal that an outbreak *may* be occurring, indicating a need for further evidence or triggering an investigation by public health officials (i.e., the CDC or a local or state public health department).

More recently, the purpose of syndromic surveillance has been expanded to include using “existing health data in real time to provide immediate analysis and feedback to those charged with investigation and follow-up of potential outbreaks.” [6] This broader focus on *electronic biosurveillance* includes both early event detection and *situational awareness*. [3] Situational awareness is the real-time analysis and display of health data to monitor the location, magnitude, and spread of an outbreak, as well as the availability and application of public health and medical resources in response to the outbreak. As Bravata et al. [7] said, “...an essential component of preparations for illnesses and syndromes potentially related to bioterrorism includes the deployment of surveillance systems that can rapidly detect *and monitor* [emphasis added] the course of an outbreak and thus minimize associated morbidity and mortality.”

Syndromic surveillance is one form of *public health surveillance*, which is the “ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health-related

event for use in public health action to reduce morbidity and mortality and to improve health.” [8] In terms of bioterrorism, syndromic surveillance uses medical and health-related data, which is distinct from other types of detection methods such as air or water sampling for pathogens (e.g., the BioWatch program [9] [10]). Some syndromic surveillance systems may be useful for epidemic monitoring, though such monitoring can be and is currently conducted via a variety of methods.

Syndromic surveillance is an emerging tool in epidemiology – the study of the distribution, determinants, and occurrence of disease and health-related conditions in populations. [11, p. 1] However, there are a number of characteristics that distinguish syndromic surveillance from public health surveillance and epidemiology as they have traditionally been practiced. For example, syndromic surveillance uses non-specific health-related data. Traditional notifiable disease reporting is based on suspected or confirmed cases. Syndromic surveillance systems are intended to actively search for evidence of possible outbreaks, where the type of outbreak is not specified in advance. In conventional public health surveillance, it is unusual to initiate active surveillance without a suspected outbreak.

Traditional epidemiological methods tend to focus on retrospective studies initiated to address a specific population risk event. In comparison, syndromic surveillance systems designed for early event detection do not have a specific risk event focus. This lack of a specific focus leads to difficult questions about how to aggregate data – syndromically, spatially, and temporally – and to questions of what types of models and detection methods are most appropriate. Given the data is usually in-hand, in classic retrospective epidemiological studies these types of questions can be directly resolved. Yet, conducting a retrospective study to try to understand the cause of a disease or outbreak can still be quite challenging. Implementing and operating a system intended to effectively detect a wide variety of possible outbreaks that have not yet occurred, from naturally occurring diseases to bioterrorism attacks, is a daunting task.

## **Surveillance Systems**

The CDC and many state and local health departments around the United States are actively developing and fielding electronic biosurveillance systems, such as the CDC’s BioSense system. Sosin [12] states that approximately 100 state and local health jurisdictions were conducting some form of syndromic surveillance in 2003. In 2004, Bravata et al. [7] conducted a systematic review of the publicly available literature and various websites from which they identified 115 surveillance systems, of which they found 29 that were designed specifically for detecting bioterrorism.

### ***Examples of Systems Currently in Use***

Below are brief descriptions of three surveillance systems chosen to illustrate large-scale systems currently in operation. The first two are true systems, in the sense that they are comprised of both dedicated computer hardware and software. The third is more properly described as a set of

software programs that can be freely downloaded and implemented by any public health organization.

- **BioSense** ([www.cdc.gov/biosense](http://www.cdc.gov/biosense)). BioSense is a federally directed collaborative effort between the CDC and the Department of Homeland Security. [13] Begun in 2003, BioSense is intended to be a United States-wide electronic biosurveillance system that initially used Department of Defense and Department of Veterans Affairs outpatient data along with medical laboratory test results from a nationwide commercial laboratory. In 2006, BioSense began incorporating data from civilian hospitals as well. The primary objective of BioSense is to “expedite event recognition and response coordination among federal, state, and local public health and healthcare organizations.” [14] [15]
- **ESSENCE** ([www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp](http://www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp)). ESSENCE is an acronym for *Electronic Surveillance System for the Early Notification of Community-based Epidemics*. Developed by the Department of Defense in 1999, ESSENCE IV now monitors for infectious disease outbreaks at more than 300 military treatment facilities worldwide on a daily basis using data from patient visits to the facilities and pharmacy data. For the Washington DC area, ESSENCE II monitors military and civilian outpatient visit data as well as over-the-counter pharmacy sales and school absenteeism. [16] [17] [18] Though ESSENCE has gone through a number of iterations, the subsequent versions have been strongly influenced by the original military project. [19]
- **EARS** ([www.bt.cdc.gov/surveillance/ears/](http://www.bt.cdc.gov/surveillance/ears/)). EARS is an acronym for *Early Aberration Reporting System*. Developed by the Centers for Disease Control and Prevention, EARS was designed for monitoring for bioterrorism during large-scale events that often have little or no baseline data (i.e., as a short-term *drop-in* surveillance method). [20] For example, the EARS system was used in the aftermath of Hurricane Katrina to monitor communicable diseases in Louisiana [21] for syndromic surveillance at the 2001 Super Bowl and World Series, as well as at the Democratic National Convention in 2000. [22] Though developed as a drop-in surveillance method, EARS is now being used on an on-going basis in many syndromic surveillance systems.

Descriptions of some state and local syndromic surveillance systems (as well as other information) can be found in [23] and [24].

### **Components of Electronic Biosurveillance Systems**

Electronic biosurveillance systems are comprised of a series of functional components. Expanding on Rolka [25], an ideal system will contain all of the components listed below.

- The original data, of which access is gained only after appropriately addressing legal and regulatory requirements, as well as personal privacy and proprietary issues.
- Computer hardware and information technology for (near) real-time assembly, recording, transfer, and preprocessing of data.
- Data management subject matter experts, software, and techniques for processing incoming data into analytic databases, as well as processes and procedures for managing and maintaining these databases.
- Statistical algorithms to analyze the data for possible outbreaks over space and time that are of sufficient sensitivity to provide signals within an actionable timeframe while simultaneously limiting false positive signals to a tolerable level.
- Public health experts with sufficient statistical expertise that can appropriately choose and apply the algorithms most relevant to their jurisdiction and appropriately interpret the signals when they occur.
- Data display and query software, as well as the necessary underlying data, that facilitates rapid and easy investigation and adjudication of signals by public health experts, including the ability to “trace back” from signal to likely source.
- Other data displays, combined with decision support and communication tools, to support situational awareness during an outbreak to facilitate effective and efficient public health response.

Mandl et al. [26], in *Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience*, provides a detailed discussion of what is required and guidance about how to implement syndromic surveillance systems.

### **Quantitative Methods Used in Biosurveillance Systems for Early Event Detection**

Biosurveillance systems use a variety of temporal and spatial methods for early event detection. As is discussed in more detail below, most of these have been adapted from other fields and applications. Buckeridge et al. [27] provides a useful classification of common surveillance scenarios and a mapping of some early event detection methods to those scenarios.

Situational awareness is an emerging concept in electronic biosurveillance and, as such, specific methods for assessing and displaying relevant information has not yet been incorporated into the systems. Currently situational awareness is limited to displaying the spatial distribution of data via geographic information system (GIS) software. (See, for example, Li et al. [28].)

### Temporal Methods

Most syndromic surveillance systems apply variants of the standard univariate statistical process control (SPC) methods: Shewhart, cumulative sum (CUSUM), and/or exponentially weighted moving average (EWMA) charts. Woodall [29] provides a comprehensive overview of the application of control charts to health surveillance. Montgomery [30] is an excellent introduction to these methods in a statistical process control setting. Shmueli and Fienberg [31] and Shmueli [32] give a review of these and other methods potentially applicable to early event detection in a biosurveillance setting.

The challenge in applying these methods is that syndromic surveillance generally violates classical SPC assumptions, particularly the assumptions of normality and independent and identically distributed observations. Biosurveillance data is generally autocorrelated and frequently has seasonal periodicities. Further, given the goal of quick detection, methods are usually run on individual observations for with an assumption of normality generally does not apply.

In spite of this, the standard SPC methods are sometimes applied with little modification (see, for example, Fricker [33] or Stoto et al. [34]) and in some cases the methods are modified to attempt to account for the autocorrelation. For example, EARS [20] applies variants of the Shewhart chart (see equations (1) and (2) below) and a cumulative method motivated by the CUSUM (see equation (3) below). These methods use various moving windows of data to estimate the process mean and standard deviation [22] and they are intended to be used when little historic (“baseline”) information is available.

The EARS methods are called “C1,” “C2,” and “C3” and are defined as follows. [35] [36] Let  $X(t)$  be an observation for period  $t$ , for example the number of individuals presenting to a particular hospital with a specific syndrome on day  $t$ . The C1 method calculates the statistic  $C_1(t)$  for day  $t$  as

$$C_1(t) = \frac{X(t) - \bar{X}_1(t)}{s_1(t)} \quad (1)$$

where  $\bar{X}_1(t) = \sum_{i=t-7}^{t-1} X(i) / 7$  and  $s_1(t) = \sqrt{\sum_{i=t-7}^{t-1} (X(i) - \bar{X}_1(i))^2 / 6}$ . It signals an alarm at time  $t$  when the  $C_1$  statistic exceeds a *threshold* which is fixed at three standard deviations above the moving sample average:  $C_1(t) > 3$ .

The C2 method is similar to the C1 method, but incorporates a two-day lag in the mean and standard deviation calculations. It calculates

$$C_2(t) = \frac{X(t) - \bar{X}_3(t)}{s_3(t)}, \quad (2)$$

where  $\bar{X}_3(t) = \sum_{i=t-3}^{t-9} X(i) / 7$  and  $s_3(t) = \sqrt{\sum_{i=t-3}^{t-9} (X(i) - \bar{X}_1(i))^2 / 6}$ , and signals an alarm when  $C_2(t) > 3$ .

Finally, the C3 method uses the C2 statistics from the past three days, calculating the statistic  $C_3(t)$  for day  $t$  as

$$C_3(t) = \sum_{i=t}^{t-2} \max[0, C_2(i) - 1]. \quad (3)$$

It signals an alarm when  $C_3(t) > 2$ .

BioSense originally implemented the C1, C2, and C3 methods, but has since modified the C2. Calling the new method “W2,” it calculates the mean and standard deviation separately for weekdays and weekends (using the relevant last seven days with a two-day lag). Following the suggestion of Buckeridge et al. [27], it uses an empiric method to define the threshold and users of the system can select the threshold as an option. [37]

To date, multivariate SPC methods have not been incorporated into operational electronic biosurveillance systems. Some research has been conducted to define and evaluate directional multivariate methods, including Fricker [33], Stoto et al. [34], and Joner et al. [38]. Whether or not multivariate methods provide additional sensitivity to detect outbreaks compared to the current practice of using multiple simultaneous univariate methods has yet to be conclusively demonstrated. Various methods have also been proposed to combine and/or adjust for the application of multiple univariate methods in multivariate settings. See the discussion in Rolka et al. [39] about parallel and consensus monitoring methods and Stoto et al. [34] and Stoto, Schonlau, and Mariano [40] for a discussion and an evaluation of some of these methods.

Regression and time series methods have also been proposed to explicitly model, and hence account for, seasonality. The basic idea is to model the disease incidence process, perhaps including terms in the model for annual seasonal variations, monthly variations, and even day of the week and holiday variations. The model is then used to either: (1) predict the expected number of events and the difference between the expected number and observed number is monitored for excessive deviations or, (2) model and then remove the explainable/known effects (sometimes called *preconditioning*) and then the residuals are monitored using traditional SPC methods. This is consistent with recommendations by Montgomery [30] for autocorrelated data.

Examples in the literature include Brillman et al. [41], who apply the CUSUM to the prediction errors, the CDC’s cyclical regression models discussed in Hutwagner et al. [22], log-linear regression models in Farrington et al. [42], and time series models in Reis and Mandl [43]. See Shmueli [32] for additional discussion of the use of regression and time series methods for syndromic surveillance and Burkom et al. [44] for a comparison of two regression-based methods

and an exponential smoothing method applied to biosurveillance forecasting. Also see Lotze et al. [45] for a detailed discussion of preconditioning applied to syndromic surveillance data.

Other temporal methods that have been proposed or are in use for syndromic surveillance include: wavelets (see Goldberg et al. [46], Zhang et al. [47], Shmueli [48], and the discussion in Shmueli [32]); Bayesian networks (see Wong et al. [49], Rolka et al. [39]); hidden Markov models (see Le Strat and Carrat [50]); Bayesian dynamic models (see, for example, Sebastiani et al. [51]), and rule-based methods (see, for example, Wong [52]).

### **Spatial and Spatio-temporal Methods**

Kleinman et al. [53] and Lazarus et al. [54] proposed a generalized linear mixed model (GLMM) to simultaneously monitor disease counts over time in a region divided into smaller sub-areas (zip codes). It is statistically attractive because it uses information across the entire region while appropriately adjusting for the smaller areas.

As described in Kleinman, et al. [53], there are two forms of the model depending on whether individual data and covariates are available versus aggregated counts and covariates by zip code. In the former case, the model is

$$E(y_{ijt} | b_i) = p_{ijt} \text{ and } \text{logit}(p_{ijt}) = \mathbf{x}_{ijt}\boldsymbol{\beta} + b_i, \quad (4)$$

where  $y_{ijt}$  is an indicator for whether or not person  $j$  in area  $i$  is a case on day  $t$ ,  $p_{ijt}$  is the probability he or she is a case,  $\mathbf{x}_{ijt}$  is a vector of observed covariates on person  $j$  and/or area  $i$  over time up to and including day  $t$ ,  $\boldsymbol{\beta}$  is a vector of fixed effects, and  $b_i$  is a random effect for area  $i$ . When no individual level covariate information is available, the most likely situation, the model is

$$E(y_{it} | b_i) = n_{it} p_{it} \text{ and } \text{logit}(p_{it}) = \mathbf{x}_{it}\boldsymbol{\beta} + b_i, \quad (5)$$

where  $y_{it} = \sum_{j=1}^{n_{it}} y_{ijt}$ . In this model,  $p_{it}$  can be thought of as the probability that each individual in area  $i$  will be a case on day  $t$ .

Having fit the model in equation (5),  $z$  cases are observed on day  $t+1$ . The rarity of the observed count is assessed by calculating

$$\Pr(Z \geq z \text{ cases}) = 1 - \sum_{k=1}^{z-1} \binom{n_{it}}{k} \hat{p}_{it}^k (1 - \hat{p}_{it})^{n_{it}-k},$$

where  $\hat{p}_{it}$  is calculated from the estimated coefficients in the usual way for logistic regression and  $1/(\hat{p}_{it} \times \# \text{ tests conducted})$  is proposed as the *recurrence interval*: the number of time periods for which the expected number of counts of  $z$  or more cases is one (Woodall et al. [55]). Waller [56] recommends an alternate calculation for the recurrence interval and Woodall et al. [55] take issue with both the use of and recommended calculations for the recurrence interval.

The Small Area Regression and Testing (“SMART”) method in BioSense (see the BioSense User Guide [57]) is based on the Kleinman et al. [53] and Lazarus et al. [54] GLMM approach. However, as implemented in BioSense it only uses spatial information to bin data into separate time series, the output of which are subsequently combined using a Bonferroni correction. Hence, the BioSense SMART method is properly classified as a temporal method. [19]

The most commonly used spatial method is the scan statistic, particularly as implemented in the SaTScan software ([www.satscan.org](http://www.satscan.org)). Originally developed to retrospectively identify disease clusters (see Kulldorff [58]), the method is now regularly used prospectively in electronic biosurveillance systems (see Kulldorff [59]). For example, it was used as part of a drop-in syndromic surveillance system in New York City after the 9/11 attack (Ackelsberg et al. [60]). While it has been studied by the BioSense program, it has not yet been implemented in the BioSense system interface (see Bradley [61] and the BioSense User Guide [57]).

The basic idea in SaTScan is to count the number of cases that occur in a cylinder, where the circle is the geographic base and the height of the cylinder corresponds to time. The cylinder is passed over space, varying the radius of the circle (up to a maximum radius that includes 50 percent of the monitored population) and the height of the cylinder and counts of cases for those geographic regions whose centroids fall within the circle for the period of time specified by the height of the cylinder are summed. When used for prospective biosurveillance, the start date of the height of the cylinder is varied but the end date is fixed at the most current time period. Conditioning on the expected spatial distribution of observations, SaTScan reports the most likely cluster (in both space and time) and its  $p$  value.

Though widely used, some aspects of the prospective application of the SaTScan methodology have been questioned, particularly the use of recurrence intervals and performance comparisons between SaTScan and other methods. See Woodall et al. [55] for further details. Also, see Kulldorff [59] for other methods for disease mapping and for testing whether an observed pattern of disease is due to chance.

Olson et al. [62] and Forsberg et al. [63] take an alternate approach to assessing possible disease clusters using M-statistics based on the distribution of pairwise distances between cases. That is, let  $\mathbf{X} = \{X_1, \dots, X_n\}$  represent the locations of  $n$  cases on the plane and let  $\mathbf{d} = \{d_1, \dots, d_{\binom{n}{2}}\}$  be the  $\binom{n}{2}$  interpoint distances, then the M-statistic is

$$M = (\mathbf{o} - \mathbf{e})^T \hat{S}^{-1} (\mathbf{o} - \mathbf{e}), \quad (6)$$

where  $\mathbf{o}$  and  $\mathbf{e}$  are vectors of observed and expected counts of binned interpoint distances and  $\hat{S}^{-1}$  is the inverse of the estimate of the covariance matrix.

Other approaches to spatial and spatio-temporal biosurveillance methods include the Automated Epidemiologic Geotemporal Integrated Surveillance System (AEGIS) by Olson et al. [64] and the application of CUSUM methods to the spatial distribution of cases (see Rogerson and Yamada [65]). See Lawson and Kleinman [66] for additional exposition and methods, and Mandl et al. [26] for further discussion on spatial and spatio-temporal modeling issues. For spatial methods with application to more traditional public health data and problems, see Waller and Gotway [67].

### **Issues and Challenges in Biosurveillance**

A number of papers discuss the various issues, challenges, and important research needs associated with effective implementation and operation of electronic biosurveillance systems. These include Sosin [1], Fricker and Rolka [3], Bravata et al. [7], Rolka [10], Shmueli [32], Stoto et al. [34], and Buehler et al. [68]. Research challenges span many disciplines and problems, from legal and regulatory challenges that must be resolved to gain access to data, to the technological challenges of designing the computer hardware and software for collecting and assembling the data, to the ethical and procedural issues for managing and safeguarding the data, to the quantitative challenges of analyzing the data for potential outbreaks and displaying the data to enhance situational awareness, to the managerial challenges of effectively assembling and operating the entire system.

Much of the controversy surrounding syndromic surveillance stems from the initial focus on early event detection, a use that requires a number of (as yet unproven) assumptions, including:

- Leading indicators of outbreaks exist in pre-diagnosis health-related data of sufficient strength that they are statistically detectable.
- The leading indicators occur sufficiently far in advance of clinical diagnoses so that, when found, they provide the public health community with sufficient advance notice to take action.

Reingold [69] has suggested that a compelling case for the implementation of syndromic surveillance systems has yet to be made and Stoto et al. [40] have questioned whether syndromic surveillance systems can achieve an effective early detection capability.

However, a myopic focus only on early event detection for bioterrorism in syndromic surveillance systems misses other important benefits electronic biosurveillance can provide, particularly the potential to significantly advance and modernize the practice of public health surveillance. For example, whether or not electronic biosurveillance systems prove effective at the early detection of bioterrorism, they are likely to have a significant and continuing role in the detection of seasonal and pandemic flu, as well as other naturally occurring disease outbreaks. This latter function is echoed in an Institute of Medicine report on *Microbial Threats to Health* [70]:

“[S]yndromic surveillance is likely to be increasingly helpful in the detection and monitoring of epidemics, as well as the evaluation of health care utilization for infectious diseases.”

In terms of bioterrorism, Stoto [71] states that electronic biosurveillance systems build links between public health and healthcare providers – links that could prove to be critical for consequence management should a bioterrorism attack occur. Furthermore, Sosin [1] points out that electronic biosurveillance systems can:

- Act as a safety net, should the existing avenues of detection fail to detect an outbreak, so countermeasures can be taken swiftly.
- Provide additional lead-time to public health authorities so that they can take more effective public health actions.

The safety net idea says that detection in a biosurveillance system does not necessarily have to be earlier than the first clinical diagnosis to be useful. To illustrate, a Dutch biologist conducting automated salmonella surveillance related that the surveillance system has detected outbreaks whose occurrence was somehow missed by sentinel physicians. [19] And the idea of additional lead time is that unusual indicators in an electronic biosurveillance system (not necessarily a signal from an early event algorithm) may give public health organizations time to begin organizing and marshalling resources in advance of a confirmed case and/or provide critical information about how and where to apply the resources.

## Conclusion

Electronic biosurveillance systems are being developed and implemented around the world. They are motivated by a need for improved public health surveillance, not only for bioterrorism, but also to improve detection and responsiveness to natural disease outbreaks such as avian influenza and SARS, and as such they hold great promise as public health tools.

For those interested in additional information, the International Society for Disease Surveillance (<http://syndromic.org>) is a professional society devoted to advancing the field of disease surveillance, and *Advance in Disease Surveillance* ([www.isdsjournal.org/](http://www.isdsjournal.org/)) is its journal.

## Acknowledgements

I would like to thank Howard Burkom, Lori Hutwagner, Michael Jackson, Galit Shmueli, Mike Stoto, Jerry Tokars, Lance Waller, and Bill Woodall for their helpful and thoughtful comments on an early version of this article. It has been significantly improved as a result of their input. Any errors or omissions that may remain are solely my responsibility.

## References

- [1] Sosin, D.M. (2003). Syndromic Surveillance: The Case for Skillful Investment View, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **1**, 247-253. Accessed online at <http://www.medscape.com/viewarticle/466780> on November 22, 2006.
- [2] Centers for Disease Control and Prevention (2006). Accessed at [www.cdc.gov/biosense/publichealth.htm](http://www.cdc.gov/biosense/publichealth.htm) on November 16, 2006.
- [3] Fricker, Jr., R.D., and H. Rolka (2007). Protecting Against Biological Terrorism: Statistical Issues in Electronic Biosurveillance, *Chance*, **20**, 1, 2007 (to appear).
- [4] International Foundation For Functional Gastrointestinal Disorders, <http://www.iffgd.org/GIDisorders/glossary.html>, accessed on November 21, 2006.
- [5] Centers for Disease Control and Prevention (2003). Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents dated October 23, 2003. Accessed online at <http://www.bt.cdc.gov/surveillance/syndromedef/> on November 21, 2006.
- [6] Hennings, K.J. (2004). "What is Syndromic Surveillance?", *Syndromic Surveillance: Reports from a National Conference, 2003*, *Morbidity and Mortality Weekly Report*, **53** (supplemental), Centers for Disease Control and Prevention, pp. 7-11.
- [7] Bravata, D.M., McDonald, K.M., Smithe, W.M., Rydzak, C., Szeto, H., Buckeridge, D.L., Haberland, C., and D.K. Owens (2004). Systematic review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases, *Annals of Internal Medicine*, **140**, 11, pp. 910- 922.
- [8] The Centers for Disease Control and Prevention (2001). Updated Guidelines for Evaluating Public Health Surveillance Systems, *Morbidity and Mortality Weekly Report*, **50**(RR13), pp. 1-35. Accessed online at <http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm> on November 21, 2006.
- [9] Chertoff, M. (2006) Testimony Before the Senate Committee on Homeland Security and Governmental Affairs, release date September 12, 2006. Accessed online at [http://www.dhs.gov/xnews/testimony/testimony\\_1158336548990.shtm](http://www.dhs.gov/xnews/testimony/testimony_1158336548990.shtm) on November 27, 2006.
- [10] Shea, D.A. and S.A. Lister (2003). The BioWatch Program: Detection of Bioterrorism, Congressional Research Service Report No. RL 32152, November 19, 2003, accessed online at <http://www.fas.org/sgp/crs/terror/RL32152.html> on November 27, 2006.
- [11] Gerstman, B.B. (1998). *Epidemiology Kept Simple: An Introduction to Classic and Modern Epidemiology*, John Wiley & Sons, New York, NY.
- [12] Sosin, D. (2005). Evaluation Challenges for Syndromic Surveillance - Making Incremental Progress, *Morbidity and Mortality Weekly Report*, **53** (supplemental), Centers for Disease Control and Prevention, pp. 125-129.
- [13] Council of State and Territorial Epidemiologists (2006). Questions and Answers: Public Health Preparedness and Disease Surveillance, accessed online at [http://www.hhs.gov/healthit/ahic/materials/meeting11/bio/CSTE\\_Surveillance\\_QA.doc](http://www.hhs.gov/healthit/ahic/materials/meeting11/bio/CSTE_Surveillance_QA.doc) on November 27, 2006.
- [14] Tokars, J. (2006). The BioSense Application, presentation at the 2006 PHIN Conference, Atlanta, GA, accessed online at [http://0-www.cdc.gov.mill1.sjlibrary.org/biosense/files/Jerry\\_Tokars.ppt#387,1](http://0-www.cdc.gov.mill1.sjlibrary.org/biosense/files/Jerry_Tokars.ppt#387,1), The BioSense Application on November 27, 2006.
- [15] The Centers for Disease Control and Prevention, BioSense, accessed at [www.cdc.gov/biosense/](http://www.cdc.gov/biosense/) on November 27, 2006.
- [16] Department of Defense, [www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp](http://www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp), accessed on November 27, 2006.
- [17] Lombardo, J.S., Burkom, H., and J. Pavlin (2004). ESSENCE II and the Framework for Evaluating Syndromic Surveillance Systems, *Morbidity and Mortality Weekly Report*, **53** (supplement), 159-165. Accessed online at [www.cdc.gov/mmwr/preview/mmwrhtml/su5301a30.htm](http://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a30.htm) on November 27, 2006.

- [18] Office of the Secretary of Defense, ESSENCE IV Improves Nation's Bio-Surveillance Capability, accessed online at [http://deploymentlink.osd.mil/news/jan05/news\\_20050125\\_001.shtml](http://deploymentlink.osd.mil/news/jan05/news_20050125_001.shtml) on November 27, 2006.
- [19] Personal communication with H. Burkom, December 22, 2006.
- [20] The Centers for Disease Control and Prevention (2006). Early Aberration Reporting System Website, [www.bt.cdc.gov/surveillance/ears/](http://www.bt.cdc.gov/surveillance/ears/), accessed on November 22, 2006.
- [21] Toprani, A., Ratard, R., Straif-Bourgeois, S., Sokol, T., Averhoff, F., Brady, J., Staten, D., Sullivan, M., Brooks, J.T., Rowe, A.K., Johnson, K., Vranken, P., Sergienko, E. (2006). Surveillance in Hurricane Evacuation Centers – Louisiana, September-October 2005, *Morbidity and Mortality Weekly Report*, **55**, Centers for Disease Control and Prevention, pp. 32-35.
- [22] Hutwagner, L., Thompson, W., Seeman, G.M., and Treadwell, T. (2003). The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS), *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, **80**, (No. 2, Supplement 1), 89i-96i.
- [23] The Centers for Disease Control and Prevention (2004). *Morbidity and Mortality Weekly Report*, **53** (supplement).
- [24] The Centers for Disease Control and Prevention (2006). Annotated Bibliography for Syndromic Surveillance, [www.cdc.gov/EPO/dphsi/syndromic/evaluation.htm](http://www.cdc.gov/EPO/dphsi/syndromic/evaluation.htm), accessed on November 28, 2006.
- [25] Rolka, H. (2006). Data Analysis Research Issues and Emerging Public Health Biosurveillance Directions, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, A. Wilson, G. Wilson, and D.H. Olwell, eds., Springer, New York, NY, pp. 101-107.
- [26] Mandl, K.D., Overhage, J.M., Wagner, M.W., Lober, W.B., Sebastiani, P., Mostashari, F., Pavlin, J.A., Gesteland, P.H., Treadwell, T., Koski, E., Hutwagner, L., Buckeridge, D.L., Aller, R.D., S. Grannis (2004). Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience, *The Journal of the American Medical Informatics Association*, **11**, pp. 141-150. Accessed online at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=353021> on November 28, 2006.
- [27] Buckeridge, D.L., Burkom, H., Campbell, M., Hogan, W.R., and A.W. Moore (2004). Algorithms for Rapid Outbreak Detection: A Research Synthesis, *Journal of Biomedical Informatics*, **38**, pp. 99-113.;
- [28] Li, H., Faruque, F., Williams, W. and R. Finley (2006). Real-Time Syndromic Surveillance, *ArcUser: The Magazine for ESRI Software Users*, January-March issue, pp. 17-19.
- [29] Woodall, W.H. (2006). The Use of Control Charts in Health-Care and Public-Health Surveillance, *Journal of Quality Technology*, **38**, pp.1-16. Available at [www.asq.org/pub/jqt](http://www.asq.org/pub/jqt).
- [30] Montgomery, D.C. (2001). *Introduction to Statistical Quality Control*, 4th edition, John Wiley & Sons, New York.
- [31] Shmueli, G., and S.E. Fienberg (2006). Current and Potential Statistical Methods for Monitoring Multiple Data Streams for Biosurveillance, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, A. Wilson, G. Wilson, and D.H. Olwell, eds., Springer, New York, NY, pp. 109-140.
- [32] Shmueli, G. (2006). Statistical Challenges in Modern Biosurveillance, *Technometrics*, draft dated September 18, 2006 (in submission).
- [33] Fricker, Jr., R.D. (2006). Directionally Sensitive Multivariate Statistical Process Control Methods with Application to Syndromic Surveillance, *Advances in Disease Surveillance*, **1**, 3, (to appear).
- [34] Stoto, M.A., Fricker, Jr., R.D., Jain, A., Diamond, A., Davies-Cole, J.O., Glymph, C., Kidane, G., Lum, G., Jones, L., Dehan, K., and C. Yuan (2006). Evaluating Statistical Methods for Syndromic Surveillance, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, A. Wilson, G. Wilson, and D.H. Olwell, eds., Springer, New York, NY, pp. 141-172.
- [35] Hutwagner, L., Browne, T., Seeman, G.M., and A.T. Fleischauer (2005). Comparing Aberration Detection Methods with Simulated Data, *Emerging Infectious Diseases*, **11**. Accessed online at [www.cdc.gov/ncidod/EID/vol11no02/04-0587.htm](http://www.cdc.gov/ncidod/EID/vol11no02/04-0587.htm) November 22, 2006.

- [36] EARS SAS program “AD\_DATA\_DAILY\_9.sas” by G.M. Seeman dated 6/03/02. Obtained via personal communication with L. Hutwagner, December 20, 2006.
- [37] The Centers for Disease Control and Prevention (2006). *BioSense Bulletin*, dated September 30, 2006.
- [38] Joner, M.D., Jr., W.H. Woodall, M.R. Reynolds, Jr., and R.D. Fricker, Jr. (2006). The Use of Multivariate Control Charts in Public Health Surveillance (draft).
- [39] Rolka, R., Burkom, H., Cooper, G.F., Kulldorff, M., Madigan, D., and W. Wong (2006). Issues in Applied Statistics for Public Health Bioterrorism Surveillance Using Multiple Data Streams: Research Needs, *Statistics in Medicine* (to appear).
- [40] Stoto, M.A., Schonlau, M., and L.T. Mariano (2004). Syndromic Surveillance: Is it Worth the Effort?, *Chance*, **17**, 19-24.
- [41] Brillman, J.C., Burr, T., Forslund, D., Joyce, E., Picard, R., and E. Umland (2005). Modeling Emergency Department Visit Patterns for Infectious Disease Complaints: Results and Application to Disease Surveillance, *BMC Medical Informatics and Decision Making*, **5**. Accessed online at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=555597> on November 24, 2006.
- [42] Farrington, C.P., Andrews, N.J., Beale, A.D., and M.A. Catchpole (1996). A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **159**, pp. 547-563.
- [43] Reis, B.Y., and K.D. Mandl (2003). Time Series Modeling for Syndromic Surveillance, *BMC Medical Informatics for Decision Making*, **3**. Accessed online at [www.biomedcentral.com/1472-6947/3/2](http://www.biomedcentral.com/1472-6947/3/2) on November 28, 2006.
- [44] Burkom, H.S., S.P. Murphy, and G. Shmueli (2006). Automated Time Series Forecasting for Biosurveillance (draft), in submission to *Statistics in Medicine*.
- [45] Lotze, T., S.P. Murphy, and G. Shmueli (2006). Preparing Biosurveillance Data for Classic Monitoring (draft), in submission to *Advances in Disease Surveillance*.
- [46] Goldberg, A., Shmueli, G., Caruana, R.A., and S.E. Fienberg (2002). Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-counter Medication Sales, *Proceeding of the National Academy of Sciences*, **99**, pp. 5237-5240. Accessed online at [www.pnas.org/cgi/reprint/99/8/5237](http://www.pnas.org/cgi/reprint/99/8/5237) on November 28, 2006.
- [47] Zhang, J., Tsui, F., Wagner, M., and W. Hogan (2003). Detection of Outbreaks from Time Series Data Using Wavelet Transform, *AMIA Annual Symposium Proceedings*, pp. 748-752. Accessed online at [www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=14728273](http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=14728273) on November 28, 2006.
- [48] Shmueli, G. (2005). Wavelet-based Monitoring for Modern Biosurveillance, technical report RHS-06-002, University of Maryland, Robert H. Smith School of Business. Accessed online at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=902878](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=902878) on November 28, 2006.
- [49] Wong, W., Cooper, G., Dash, D., Levander, J., Dowling, J., Hogan, W., and M. Wagner (2005). Use of Multiple Data Streams to Conduct Bayesian Biologic Surveillance, *Morbidity and Mortality Weekly Report*, **54** (supplemental), pp. 63-69. Accessed online at <http://rods.health.pitt.edu/LIBRARY/Wong.%20Use%20of%20Multiple%20Data%20Streams%20to%20Conduct%20Bayesian%20Biologic%20Surveillance.%20MMWR%2054%20Suppl%20.pdf> on November 28, 2006.
- [50] Le Strat, Y., and F. Carrat (1999). Monitoring Epidemiologic Surveillance Data Using Hidden Markov Models, *Statistics in Medicine*, **18**, pp. 3463-78. Accessed online at <http://www3.interscience.wiley.com/cgi-bin/fulltext/68502848/PDFSTART> on November 28, 2006.
- [51] Sebastiani, P., Mandl, K.D., Szolovits, P., Kohane, I.S., and M.F. Ramioli (2006). A Bayesian Dynamic Model for Influenza (with discussion), *Statistics in Medicine*, **25**, pp. 1803-1825.
- [52] Wong, W., Moore, A., Cooper, G., and M. Wagner (2003). WSARE: What’s Strange About Recent Events?. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, **80** (supplement), pp. 66i-75i. Accessed online at <http://rods.health.pitt.edu/LIBRARY/WSARE%20What's%20Strange%20%20J%20Urban%20Health%20Cooper%202003.pdf> on November 28, 2006.

- [53] Kleinman, K., Lazarus, R., and R. Platt (2004). A Generalized Mixed Model Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism, *American Journal of Epidemiology*, **159**, pp. 217-224.
- [54] Lazarus, R., Kleinman, K., Dashevsky, I., Adams, C., Kludt, P., DeMaria, Jr., A., and R. Platt (2002). Use of Automated Ambulatory-Care Encounter Records for Detection of Acute Illness Clusters, Including Potential Bioterrorism Events, *Emerging Infectious Diseases*, **8**, pp. 753-760. Accessed online at [www.medscape.com/viewarticle/440756\\_print](http://www.medscape.com/viewarticle/440756_print) on November 28, 2006.
- [55] Woodall, W.H., Marshall, J.B., Joner, M.D., Jr., Fraker, S.E., and A.G. Abdel-Salam (2006). On the Use of Scan Methods in Prospective Public Health Surveillance, to be submitted to *Journal of the Royal Statistical Society, Series A (Statistics in Society)*.
- [56] Waller, L.A. (2004). Invited Commentary: Syndromic Surveillance—Some Statistical Comments, *American Journal of Epidemiology*, **159**, pp. 225-227.
- [57] The Centers for Disease Control and Prevention (2006). BioSense User Guide, Version 2.0, March 2006. Accessed online at [http://0-www.cdc.gov.mill1.sjlibrary.org/biosense/files/CDC\\_BioSense\\_User\\_Guide\\_v2.0.pdf](http://0-www.cdc.gov.mill1.sjlibrary.org/biosense/files/CDC_BioSense_User_Guide_v2.0.pdf) on November 28, 2006.
- [58] Kulldorff, M. (1997). A Spatial Scan Statistic, *Communications in Statistics, Theory and Methods*, **26**, pp. 1481-1496. Accessed online at [www.satscan.org/papers/k-cstm1997.pdf](http://www.satscan.org/papers/k-cstm1997.pdf) on November 28, 2006.
- [59] Kulldorff, M. (2001). Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **164**, pp. 61-72. Accessed online at [www.satscan.org/papers/k-jrssa2001.pdf](http://www.satscan.org/papers/k-jrssa2001.pdf) on November 28, 2006.
- [60] Ackelsberg, J., Balter, S., Bornschelgel, K., Carubis, E., Cherry, C., Das, D., Fine, A., Karpati, A., Layton, M., Mostashari, F., Nivin, B., Reddy, V., Weiss, D., Hutwagner, L., Seeman, G.M., McQuiston, J., Treadwell, T., and J. Rhodes (2002). Syndromic Surveillance for Bioterrorism Following the Attacks on the World Trade Center - New York City, 2001, *Morbidity and Mortality Weekly Report*, **51** (Special Issue), Centers for Disease Control and Prevention, pp. 13-15.
- [61] Bradley, C. (2005). Visualizing and Monitoring Data in BioSense Using SaTScan, presentation at the 2005 Syndromic Surveillance Conference, accessed online at [http://www.cdc.gov/biosense/files/SaTScan\\_Presentation\\_2005.ppt#501,1,Colleen Bradley, MSPH Syndromic Surveillance Conference, 2005](http://www.cdc.gov/biosense/files/SaTScan_Presentation_2005.ppt#501,1,Colleen%20Bradley,%20MSPH%20Syndromic%20Surveillance%20Conference,%202005) on November 28, 2006.
- [62] Olson, K.L., Bonetti, M., Pagano, M., and K.D. Mandl (2005). Real Time Spatial Cluster Detection Using Interpoint Distances Among Precise Patient Locations, *BMC Medical Informatics and Decision Making*, **5**. Accessed online at <http://www.biomedcentral.com/1472-6947/5/19> on December 4, 2006.
- [63] Forsberg, L., Jeffery, C., Ozonoff, A., and M. Pagano (2006). A Spatiotemporal Analysis of Syndromic Data for Biosurveillance, *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, A. Wilson, G. Wilson, and D.H. Olwell, eds., Springer, New York, NY, pp. 173-191.
- [64] Olson, K.L., Bonetti, M., Pagano, M., and K.D. Mandl (2002). Enhanced Power to Detect Bioterrorism with Spatial Clustering (abstract), *Pediatric Research*, **51**, p. 94.
- [65] Rogerson, P.A., and I. Yamada (2004). Monitoring Change in Spatial Patterns of Disease: Comparing Univariate and Multivariate Cumulative Sum Approaches, *Statistics in Medicine*, **23**, pp. 2195-2214.
- [66] Lawson, A.B., and K. Kleinman (eds.) (2005). *Spatial and Syndromic Surveillance for Public Health*, John Wiley & Sons.
- [67] Waller, L.A., and C.A. Gotway (2004). *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons.
- [68] Buehler, J.W., Berkelman, R.L., Hartley, D.M., and C.J. Peters (2003). Syndromic Surveillance and Bioterrorism-related Epidemics, *Emerging Infectious Diseases*, **9**, 1197-1204.
- [69] Reingold, A. (2003). If Syndromic Surveillance is the Answer, What is the Question?, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **1**, 1-5. Accessed online at <http://www.medscape.com/viewarticle/458654> on November 22, 2006.

[70] Smolinski, M.S., Hamburg, M.A., Lederberg, J., eds., *Microbial Threats to Health: Emergence, Detection, and Response*, National Academies Press.

[71] Stoto, M.A., Syndromic Surveillance in Public Health Practice, presentation to Institute of Medicine Forum on Microbial Threats, December 12, 2006.

DRAFT