

CHAPTER 1

**SURVEY OF EARLY WARNING SYSTEMS FOR
ENVIRONMENTAL AND PUBLIC HEALTH APPLICATIONS**

Chung-Sheng Li

IBM T. J. Watson Research Center, PO Box 704, Yorktown Heights, NY 10598

1. Introduction

Advances in sensors and actuators technologies and real-time analytics recently have accelerated the adoption of real-time or near real-time monitoring and alert generation for environmental and public health related activities.

Monitoring of environmental related activities include the use of remote sensing (satellite imaging and aerial photo) for tracking the impacts due to global climate change (e.g. detection of thinning polar ice as evidence of global warming, deforestation in the Amazon region), tracking the impact of natural and man-made disasters (e.g. earthquakes, tsunami, volcano eruptions, floods, and hurricane), forest fire, and pollution (including pollution of air, land, and water).

Early warnings for human disease or disorder at the individual level often rely on various symptoms or biomarkers. For example, retina scan has been developed as a tool for early diagnosis of diabetic retinopathy [Denninghoff 2000]. CA 125, a protein found in blood, has been used as the biomarker for screening early stages of ovarian cancer. The level of Prostate Specific Antigen (PSA) has been used for early detection of prostate cancer (usually in conjunction with a digital rectal exam). National Cancer Institute of NIH has sponsored the forming of the Early Detection Research Network (EDRN

<http://www3.cancer.gov/prevention/cbrg/edrn/>) program since 1999. This comprehensive effort includes the development and validation of various biomarkers for detecting colon, liver, kidney, prostate, bladder, breast, ovary, lung, and pancreatic cancer.

Health and infectious disease surveillance for public health purposes have existed for many decades. Currently, World Health Organization (WHO) has set up a Global Outbreak Alert & Response Network (GOARN) to provide surveillance and response for communicable diseases around the world (<http://www.who.int/csr/outbreaknetwork/en/>). Examples of the diseases tracked by GOARN include anthrax, influenza, Ebola hemorrhagic fever, plague, smallpox, and SARS. The alert and response operations provided include epidemic intelligence, event verification, information management and dissemination, real time alert, coordinated rapid outbreak response, and outbreak response logistics. For the United States, the National Center for Infectious Diseases (NCID) at CDC has set up 30+ surveillance networks (see Section 2) to track individual infectious diseases or infectious disease groups.

Monitoring disease outbreaks for public health purposes based on environmental epidemiology has been demonstrated for a number of *vector-borne infectious diseases* such as Hantavirus Pulmonary Syndrome (HPS), malaria, Lyme disease, West Nile virus (WNV), and Dengue fever [Glass 2001a, Glass 2001b, Glass 2002, Klein 2001a, Klein 2001b]. In these cases, environmental effects such as the change of ground moisture, ground temperature, and vegetation due to El Niño and other climate changes facilitate the population change of disease hosts (such as mosquitoes and rodents). These changes can then cause increased risk of diseases in the human population.

Recently, a new trend to track subtle human behavior changes due to disease outbreak has emerged to provide advanced warnings before significant casualties registered from clinical sources. This approach is also referred to in the literature as *syndromic surveillance* (or early detection of disease outbreaks, pre-diagnosis surveillance, non-traditional surveillance, enhanced surveillance, non-traditional surveillance, and disease early warning systems). This approach has received substantial interests and enthusiasm during the past few years, especially after Sept.

11, 2001 [Buehler 2003, Duchin 2003, Goodwin2004, Mostashari 2003, Pavlin 2003, Sosin 2003, and Wagner 2001].

In this chapter, we surveyed the most widely adopted disease surveillance techniques, early warning systems, and detection approaches in terms of problem domain and data sources. This survey intends to facilitate the clarification and discovery of interdependency among problem domains, data sources, and detection methods and thus facilitate sharing of data sources and detection methods across multiple problem domains, when appropriate.

2. Disease Surveillance

Health and infectious disease surveillance for public health purposes have existed for many decades. The following is a subset of the 30+ national surveillance systems coordinated by the National Center of Infectious Disease (NCID) at CDC (http://www.cdc.gov/ncidod/osr/site/surv_resources/surv_sys.htm).

Mortality Reporting System: from 122 cities & metropolitan areas, compiled by the CDC epidemiology program office;

Active Bacterial Core Surveillance (ABCs): at ten emerging infectious program sites such as California, Colorado, Connecticut, Georgia, etc.;

BaCon Study: on assessing the frequency of blood component bacterial contamination associated with transfusion;

Border Infectious Disease Surveillance Project (BIDS): focusing on hepatitis and febrile-rash illness along the US-Mexico border;

Dialysis Surveillance Network (DSN): focus on tracking vascular access infections and other bacterial infections in hemodialysis patients;

Electronic Foodborne Outbreak Investigation and Reporting System (EFORS): used by 50 states to report data about Foodborne Outbreaks on a daily basis;

EMERGENCY ID NET: focus on emerging infectious disease based on 11 university affiliated urban hospital emergency departments;

Foodborne Diseases Active Surveillance Network (FoodNet): consists of active surveillance for foodborne diseases and related epidemiologic studies. This is a collaboration among CDC, the ten emerging infectious program sites (EIPs), the USDA, and FDA;

Global Emerging Infectious Sentinel Network (GeoSentinel): consists of travel/tropical medicine clinics around the world that monitor geographic and temporal trends in mobility;

National Notifiable Diseases Surveillance System (NNDSS): The Epidemiology Program Office (EPO) at CDC collect, compile, and publish reports of disease considered as notifiable at the national level;

National West Nile Virus Surveillance System: Since 2000, 48 states and 4 cities started to collect data about wild birds, sentinel chicken flocks, human cases, veterinary cases, and mosquito surveillance;

Public Health Laboratory Information System (PHLIS): Collects data on cases/isolates of specific notifiable diseases from every state within the United States.

United States Influenza Sentinel Physicians Surveillance Network: 250+ physicians around the country report each week the number of patients seen and the total number with flu-like symptoms;

Waterborne-Disease Outbreak Surveillance System: includes data regarding outbreaks associated with drinking water and recreational water.

The US DoD Global Emerging Infectious Surveillance System (DoD-GEIS): has its own set of surveillance network within the States as well as internationally (<http://www.geis.fhp.osd.mil/>).

A number of syndromic surveillance prototypes or deployments have been developed recently:

BioSense: a national initiative at US (coordinated by CDC) to establish near real-time electronic transmission of data to local, state, and federal public health agencies from national, regional, and local health data by accessing and analyzing diagnostic and prediagnostic health data.

RODS (Real Time Outbreak and Disease Surveillance <http://rods.health.pitt.edu/>): a joint effort between University of Pittsburgh and CMU. This effort includes an open source version of RODS software (since 2003) and National Retail Data Monitor (NRDM) which monitors the sales of over-the-counter healthcare products in 20,000 participating pharmacies in the states.

The New York City Department of Health and Mental Hygiene [Greenko 2003] has established a syndromic surveillance system since 2001 that monitors emergency department visits to provide early detection of disease outbreaks. The key signs and symptoms analyzed include respiratory problems, fever, diarrhea, and vomiting.

Concerns do exist for the effectiveness and general applicability of syndromic surveillance [Stoto 2004]. In particular, the size and timing of an outbreak for which syndromic surveillance gives an early advantage may be limited. In a case involving hundreds or thousands of simultaneous infections, no special detection methods would be needed. Conversely, even the best syndromic surveillance system might not work in a case involving only a few individuals such as the anthrax episodes of 2001 in the United States. Consequently, the size of an outbreak in which syndromic surveillance can provide added value (on top and beyond the traditional infectious disease surveillance) is probably on the order of tens to several hundred people.

3. Reference Architecture for Model Extraction

Figure 1 illustrates the general reference architecture of an end-to-end surveillance and response system. The surveillance module is responsible for collecting multi-modal data from available data sources, and conduct preliminary analysis based on one or more data sources. The surveillance module works closely with the simulation and modeling module to extract the model from these data sources. The primary responsibility for the simulation and modeling module is to determine whether an anomaly has already happened or is likely to happen in the future. Furthermore, it is also responsible for interacting with the decision maker to provide decision support on multiple what-if scenarios. The decision maker can then use the command and control module to launch appropriate actions (such as deploy antibiotics, applying ring vaccination, and quarantine) based on the decisions.

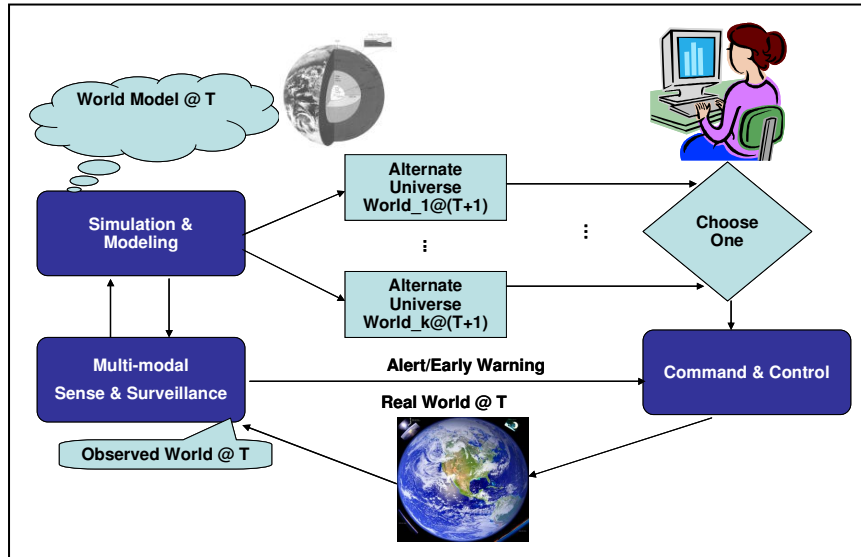


Figure 1: Reference architecture of an end-to-end surveillance & response system.

Surveillance and activity monitoring in environmental and public health domain often involves (1) collecting data from clinical data (patient records, billing information, lab results, etc.) as well as both hardware and/or software sensors which generate multi-modal multi-scale spatio-temporal data, (2) cleaning and preprocessing the data to remove gaps and inconsistencies, if any, and (3) applying anomaly detection methodologies to generate predictions and early warnings. Predictions are often based on the environmental and public health threat models.

As an example, we might subconsciously or consciously adjust our diet when we feel ill (such as drinking more water, juice, and have more rest). If the symptoms become more severe, we might seek over-the-counter (OTC) medicine, and miss classes/work. In many cases, we might go to work late or leave for home early. We might also experience subtle change of our behavior at work. When the symptoms continue, we might seek help from physicians. We usually have to make appointments with the physicians (e.g. through phone or through web) first. Physician visits often result in prescription medicines and thus visits to the pharmacies.

This behavior model suggests that we could potentially observe the progression of a disease outbreak within a population at multiple touch points by this model – such as sales of drinking water, sewage generation, OTC drug sales, absentee information from school and work, phone records to physician office, and 911 calls. Some of these data may be collected before visits to the physician or hospital have actually happened, and thus enable earlier detection. Data can also be collected during and after visits to physicians have been made. These data could include diagnosis made by the physician (including chief complaints), lab tests, and drug prescriptions. Similar models can also be established for pollution, non-infectious diseases, and other natural disasters.

We assume that the real world phenomenon, whether it is a disease outbreak or an environmental event, will induce noticeable changes in human, animal, or other environmental behaviors. The behavioral change can be potentially captured from various hardware and/or software sensors as well as traditional business transactions. The challenge for early warning is to fuse those appropriate multi-modal and multi-scale data sources to extract the behavior model of human or environment and compared to the behavior under normal conditions (if known) and declare the existence of anomaly when the behavior is believed to be abnormal.

Figure 2 shows that the detection and early warning approaches in environmental and public health application can be characterized along three dimensions: problem domain, data sources, and detection methodologies, each of which will be further elaborated in the following three sections. These three dimensions are also the key aspects for materializing the reference architecture shown in Fig. 1 for any specific application. These three dimensions are often strongly correlated – as a given problem domain (such as detecting influenza outbreak) will dictate the available data sources and the most suitable detection methodologies.

Figure 3 shows the reference architecture of a Health Activity Monitoring system that materializes the multi-modal sense/surveillance and modeling modules in Fig. 1. Data from various sources (both traditional and non-traditional data) are streaming through the messaging

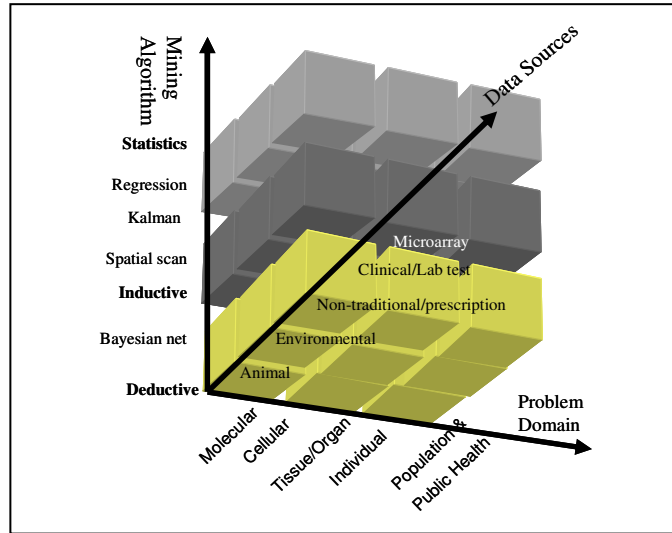


Figure 2: Taxonomy of data mining issues.

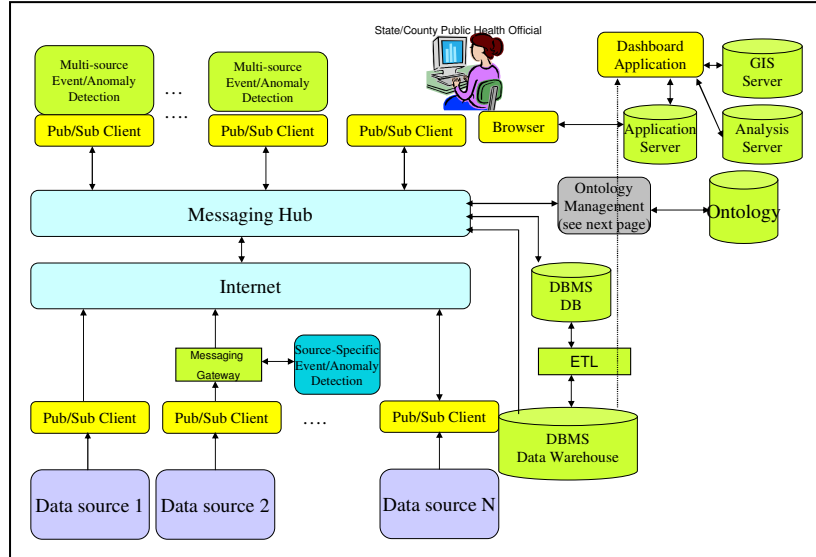


Figure 3: Reference Architecture for Health Activity Monitoring.

hub to anomaly detectors. A messaging hub is responsible for delivering the messages and events from the publishers to the intended subscribers. The publisher/subscriber clients allow the data to be streamed through for any clients which subscribe to the specific data sources. Message gateways can also be attached to the pub/sub clients for source-specific data analysis before sending to the messaging hub.

Each anomaly detector can subscribe to one or more data sources, and publish alerts for other anomaly detectors. The incoming data are also sent into an operational database. An ETL (extract-transform-load) module can periodically pull the data from the operational database into a data warehouse for historical analysis. The analysis sometimes is conducted with the help from a GIS server that stores the geographic information. Additional statistical analysis can be performed on the historical data by an analysis server. The public health personnel can access alerts published in real time by the anomaly detectors as well as use the analysis server and data warehouse to analyze the historical data.

Ontology and metadata (as shown in Fig. 3) for data sources can often be used to assist in the interpretation of data at the appropriate abstraction level. In the disease classification area, ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. As an example, diseases of the respiratory system are coded with 460-519. New codes may be assigned to newly discovered diseases – such as Severe Acute Respiratory Syndrome - SARS (480.3) has just been added under the category of viral pneumonia (480) since 2003.

4. Problem Domain

The problem domains can be broadly classified according to the scale of impact:

Non-infectious disease: such as cancer, cardiovascular disease, dementia, and diabetes;

Infectious diseases: which include communicable diseases (such as flu and smallpox, which can spread from human to human), vector-born

diseases which are carried by rodents or mosquitoes (such as West Nile encephalitis, Dengue fever, malaria, and Hantavirus Pulmonary Disease).

Note that non-infectious diseases of a population can potentially have a common root cause, such as caused by the same type of pollutant, similar work condition, etc. Each of the problem domains can potentially be associated with the progression of symptoms – such as sore throat, running nose, coughing, fever, vomiting, dizziness, etc. Most of the diagnoses done before the confirmation of lab tests are based on apparent symptoms – and hence the term - syndromic surveillance.

The problem domain usually determines the data sources that will be meaningful for early detection as well as which detection methods can be applied. As an example, sewage data will be more useful for GI (gastrointestinal) related disorders, while pollutants (from Environmental Protection Agency - EPA) will be more useful for respiratory diseases or disorders.

5. Data Sources

Data sources for early detection can be categorized into:

Clinically related data – the data acquired in a clinical setting such as patient records (such as the reference information model defined by HL7), insurance claims (e.g. CMS-1500 for physician visits and UB-92 for hospital visits), prescription drug claims (NCPDP), lab tests, imaging, etc. These claim forms can be converted into an electronic format - HIPAA 837 (HIPAA stands for Health Information Portability and Accountability Act) and vice versa. As an example, CMS-1500 contains information such as patient age, gender, geography, ICD-9 diagnosis codes, CPT procedure codes, date of service, physician, location of care, and payer type. NCPDP form for prescription drug contains, in addition to the basic patient data, Rx date written, Rx date filled, NDC (form/strength), quantity dispensed, days supply, specialty, payer type, and pharmacy.

Non-traditional Data - Emerging syndromic surveillance applications also start to use non-traditional (non-clinical) data sources to

observe population behavior, such as retail sales (including over-the-counter drug sales), absentee data from school and work, chief complaints from the emergency room, and 911 calls. These data sources do not necessarily have standardized formats. Additional data sources include audio/video recording of seminars (for counting cough episodes).

Auxiliary Data - Additional data sources include animal data, environmental data (ground moisture, ground temperature, and pollutants) and micro-array data. Many of these data sources are problem domain dependent.

A number of studies have been devoted to investigating various data sources, such as the text and diagnosis code of the chief complaints from emergency department [Begier 2003, Beitel 2004, Espino 2001, Greenko 2003], 911 calls [Dockrey 2002], and over-the-counter (OTC) drug sales [Goldenberg 2002].

The data sources can be further characterized by its modality:

Structured: most of the transactional records stored in a relational database belong to this category. Each field of a transaction is often well defined. This may include a portion of the clinical & patient records - such as patient name, visit date, care provider name and address, diagnosis in terms of International Classification of Disease code (or ICD-9 code).

Semi-structured: refers to data that has been tagged by mark up languages such as XML. Emerging data standards such as HL7 enable the transmission of healthcare related information in XML encoded format.

Non-structured: Patient records, lab reports, emergency room diagnosis and 911 calls could include chief complaints in free text format. Additional non-structured data include data from various instruments such as CT, PET, MRI, EKG, ultrasound and from clinical and non-clinical data such as weather maps, videos, and audio segments.

The type of data sources is often essential in determining the methodologies for detecting anomalies.

6. Detection Methods

The analytic methods are often data source dependent. Spatial scan statistics has been most widely used, such as in [Berkom 2003]. Other methods, including linear mixed models [Kleinman 2004], space-time clustering [Koch 2001], CuSum [O'Brien 1997], time-series modeling [Reis 2003], rule-based [Wong 2003], and concept-based [Buckridge 2002], have also been investigated and compared. A comprehensive comparison among various detection algorithms is conducted in [Buckridge 2005] as a result of the Darpa sponsored BioALIRT program.

The detection methods are used to determine the existence of anomalies in the problem domain from available data sources, sometimes with either explicit or implicit assumptions of models. These detection methods, as shown in Fig. 2, can be categorized into:

Inductive method: Most of the statistical methods such as those based on linear regression, Kalman filter, and spatial scan all fall into this category. Inductive methods can be further categorized as supervised and unsupervised methods. Supervised methods usually involve learning and/or training a model based on historical data. Unsupervised methods, on the other hand, do not require historical data for training. Unsupervised methods are more preferable for those rare events (such as those potentially caused by bio-terrorism) when there are insufficient historical data. It has been concluded in [Buckridge 2005] that spatial and other covariate information from disparate sources could improve the timeliness of outbreak detection. Some of the existing syndromic surveillance deployments such as the CDC BioSense and New York City have leveraged both CuSum and Spatial Scan statistics.

Deductive methods: Deductive methods usually involve rule-based reasoning and inferencing. This methodology is often used in the area of knowledge-based surveillance in which existing pertinent knowledge and data are applied in making inferences from the newly arrived data.

Hybrid inductive/deductive methods: Both inductive and deductive methods can be used simultaneously on a given problem. Deductive methods can be used to select the proper inductive method given the prior knowledge in the problem domain and available data sources. This

often results in improved performance as compared to using either inductive or deductive methods alone [Buckridge 2005].

7. Summary and Conclusion

Early warning systems for environmental and public health applications have received substantial attention during the recent past due to increased awareness of infectious diseases and availability of surveillance technologies. In this chapter, we surveyed a number of existing early warning systems in terms of their problem domains (non-infectious disease vs. infectious disease), data sources (clinical data, non-traditional data and auxiliary data), and detection approaches (inductive, deductive, and hybrid). The rapid progress in the syndromic surveillance area seems to suggest that early detection of certain infectious diseases is feasible. Nevertheless, there is substantial interdependency among problem domains, data sources, and detection methods for syndromic surveillance. A validated early warning or anomaly detection approach for a given problem area often cannot be generalized to other problem areas due to the nature of data sources. Further studies in this area will likely to be based on the ontology (or taxonomy) of the diseases, data sources, and detection approaches.

Acknowledgment

The authors would like to acknowledge the EpiSPIRE bio-surveillance team at IBM T. J. Watson Research Center. This research is sponsored in part by the Defense Advanced Research Projects Agency and managed by Air Force Research Laboratory under contract F30602-01-C-0184 and NASA/IBM CAN NCC5-305.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the Defense Advanced Research Projects Agency, Air Force Research Lab, NASA, or the United States Government.

References

- Begier, E. M. , D. Sockwell, L. M. Branch, J. O. Davies-Cole, L. H. Jones, L. Edwards, J. A. Casani, D. Blythe. (2003) The National Capitol Region's emergency department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? *Emerging Infectious Disease*, vol. 9, no. 3, pp. 393-396, March.
- Beitel, A. J. , K. L. Olson, B. Y. Reis, K. D. Mandl. (2004) "Use of Emergency Department Chief Complaint and Diagnostic Codes for Identifying Respiratory Illness in a Pediatric Population," *Pediatric Emergency Care*, vol 20, no. 6, pp. 355-360, June, 2004.
- Buckeridge, D. L. , J. K. Graham, M. J. O'Connor, M. K. Choy, S. W. Tu, M. A. Musen. (2002) Knowledge-based bioterrorism surveillance, *Proceedings AMIA Symposium 2002*: 76-80.
- Buckeridge, D. L., H. Burkom, M. Campbell, W. R. Hoganb, A. W. Moore. [2005] Algorithms for Rapid Outbreak Detection: A Research Synthesis. *Biomedical Informatics*, 2005 April, 38(2):99-113.
- Buehler, J. W., R. L. Berkelman, D. M. Hartley, C. J. Peters (2003) "Syndromic Surveillance and Bioterrorism-related epidemics," *Emerging Infectious Diseases*, 2003 Oct. vol. 9 no. 10, pp 1197-1204.
- Burkom, H. S. (2003) "Biosurveillance applying scan statistics with multiple, disparate data source," *Journal of Urban Health*, 2003, June; 80 Suppl 1:i57-i65.
- Denninghoff, Kurt R., M. H. Smith, L. Hillman. (2000) Retinal Imaging Techniques in Diabetes, *Diabetes Technology & Therapeutics*, vol. 2, no. 1, pp. 111-113.
- Dockrey, M. R. , L. J. Trigg, W. B. Lober. (2002) An Information Systems for 911 Dispatch Monitoring System and Analysis, *Proceeding of the AMIA 2002 Annual Symposium* pp. 1008.
- Duchin, J. S. (2003) Epidemiological Response to Syndromic Surveillance Signals, *Journal of Urban Health*, 2003 80:i115-i116.
- Espino, J. U. , M. M. Wagner. (2001) Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness, *Proceedings AMIA Symposium 2001*, pp164-168.
- Glass, GE, T. L. Yates, J. B. Fine, T. M. Shields, J. B. Kendall, A. G. Hope, C. A. Parmenter, C.J. Peters, T. G. Ksiazek, C.-S. Li, J. A. Patz and J. N. Mills. (2002) Satellite imagery characterizes local animal reservoir populations of Sin Nombre virus in the southwestern United States, *Proc. National Academy of Science* 99:16817-16822.
- Glass, G. E. (2001a) Public health applications of near real time weather data, *Proc. 6th Earth Sciences Information Partnership Conf.*
- Glass G. E. (2001b) Hantaviruses - Climate Impacts and Integrated Assessment, *Energy Modeling Forum*.

- Goldenberg, A., G. Shmueli, R. A. Caruana, S. E. Fienberg. (2002) "Early Statistical Detection of Anthrax outbreaks by tracking over-the-counter medication sales," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 8, pp. 5237-5240, April 16, 2002.
- Goodwin, T. , and E. Noji, (2004) Syndromic Surveillance, *European Journal of Emergency Medicine*, 2004 Feb; vol. 11, no. 1, pp 1-2.
- Greenko, J. , F. Mosgtashari, A. Fine, M. Layton. (2003) Clinical Evaluation of the Emergency Medical Services (EMS) Ambulance Dispatch-Based Syndromic Surveillance System, New York City, *Journal of Urban Health*, 2003 Un; 80 Suppl 1:i50-i56.
- Klein, S. L., A. L. Marson, A. L. Scott, GE Glass. (2001a) Sex differences in hantavirus infection are altered by neonatal hormone manipulation in Norway rats, *Soc Neuroscience*.
- Klein, S. L. , A. L. Scott, G. E. Glass. (2001b) Sex differences in hantavirus infection: interactions among hormones, genes, and immunity, *Am Physiol Soc*.
- Kleinman, K., R. Lazarus, R. Platt. (2004) "A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism," *American Journal of Epidemiology*, 2004 Feb 1; 159(30): 217-224.
- Koch, M. W. , S. A. Mckenna. (2001) "Near Real Time Surveillance Against Bioterror Attack Using Space-Time Clustering, Technical Report, (<http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2001/010820p.pdf>)
- Mostashari, F. , J. Hartman. (2003) "Syndromic Surveillance: A Local Perspective," *Journal of Urban Health*, 2003; 80 Suppl 1:i1-i7.
- O'Brien, S. J., P. Christie. (1997) Do CuSums have a role in routine communicable disease surveillance, *Public Health* , July, 11194: 255-8.
- Pavlin, J. A. (2003) Investigation of Disease Outbreaks Detected by Syndromic Surveillance Systems, *Journal of Urban Health*, 2003; 80:i107-i114.
- Sosin, D. M. (2003) Syndromic Surveillance: The case for skillful investment. Biosecurity and Bioterrorism: Biodefense strategy, *Practice and Science* 2003 vol 1, no. 4, pp.247-253.
- Stoto, M. A., M. Schonlau, and L. T. Mariano. (2004) Syndromic Surveillance: Is It Worth the Effort? *Chance*, Vol. 17, No. 1, 2004, pp. 19-24
- Reis, B. Y. and K. D. Mandle. (2003) Time Series Modeling for Syndromic Surveillance, *BMC Medical Informatics and Decision Making*, Jan 23; 3(1); 2.
- Wagner, M. M. , F. C. Tsui, J. U. Espino, V. M. Dato, D. F. Sittig, R. A. Caruana, L. F. McGinnis, D. W. Deerfield, M. J. Druzdzal, D. B. Fridsma. (2001) The emerging science of very early detection of disease outbreaks, *Journal of Public Health Management Practice*, 2001 Nov; vol 7, no. 6, pp. 51-59.
- Wong, W. , A. Moore, G. Cooper, and M. Wagner. (2003) "Rule-Based Anomaly Pattern Detection for Detecting Disease Outbreaks," *Journal of Urban health*, June 80 Suppl 1: i66-i75.