Chapter 1

INFECTIOUS DISEASE INFORMATICS: AN INTRODUCTION AND AN ANALYSIS FRAMEWORK

Syndromic surveillance is concerned with continuous monitoring of public health-related information sources and early detection of adverse disease events. In practice, syndromic surveillance systems are being increasingly adopted to meet the critical needs of effective prevention, detection, and management of infectious disease outbreaks, either naturally-occurring or caused by bioterrorism attacks. From an academic standpoint, syndromic surveillance research is by nature multidisciplinary and has been attracting significant attention in recent years. This monograph presents a comprehensive review of the state of the art of syndromic surveillance research and system development efforts from the perspective of information science and logics. On the basis of a detailed analysis of more than 50 local, state, national, and international syndromic surveillance systems and a review of about 200 academic publications, in this monograph we discuss the technical challenges, applicable approaches or solutions, and the current state of system implementation and adoption for key components of syndromic surveillance systems ranging from system architecture, data collection and sharing, data analysis, and data access and visualization. In addition, we present several case studies to compare several state-of-the-art syndromic surveillance systems. The purpose of these case studies is to illustrate the information technology-driven technical discussions in an integrated, real-world context. We also briefly touch upon critical nontechnical issues including data sharing policies, and system evaluation and adoption.

This introductory chapter briefly discusses the importance of syndromic surveillance and what we believe to be/is a unique niche this book intends to fill.
In this time of increasing concern over the deadly and costly threats of infectious diseases caused by natural disasters or bioterrorism attacks, preparation for, early detection of, and timely response to emerging infectious diseases and epidemic outbreaks are a key public health priority and are driving an emerging field of multidisciplinary research. A few recent disastrous events that threatened the public health of large populations around the world include the Severe Acute Respiratory Syndrome epidemics (SARS) originated in Asia (Li et al., 2004), the outbreak of Avian flu in East Asian countries (NBII, 2006; USDA, 2006), and the ever pending threats of bioterrorism since the anthrax attacks in October 2001 (Buehler et al., 2003; Cronin, 2005; Siegrist, 1999).

Public health surveillance has been practiced for decades and continues to be an indispensable approach for detecting emerging disease outbreaks and epidemics. Early knowledge of a disease outbreak plays an important role in improving response effectiveness (Pinner et al., 2003). Although traditional disease surveillance often relies on time-consuming laboratory diagnosis and the reporting of notifiable diseases is often slow and incomplete, a new breed of public health surveillance systems has the potential to significantly speed up detection of disease outbreaks. These new, computer-based surveillance systems offer valuable and timely information to hospitals as well as to state, local, and federal health officials (Dembek et al., 2005; Pavlin, 2003). These systems are capable of real-time or near real-time detection of serious illnesses and potential bioterrorism agent exposures, allowing for a rapid public health response. This public health surveillance approach is generally called *syndromic surveillance*, which is defined as an ongoing, systematic collection, analysis, and interpretation of “syndrome”-specific data for early detection of public health aberrations.

The rationale behind syndromic surveillance lies in the fact that specific diseases of interest can be monitored by syndromic presentations that can be shown in a timely manner such as nurse calls, medication purchases, and school or work absenteeism. In addition to early detection and reporting of monitored diseases, syndromic surveillance also provides a rich data repository and highly active communication system for situation awareness and event characterization. Multiple participants provide interconnectivity among disparate and geographically separated sources of information to facilitate a clear understanding of the evolving situation. This is of significant importance for event reporting, strategic response planning, and disaster victim tracking. Information gained from syndromic surveillance data can also guide the planning, implementation, and evaluation of long-term programs to prevent and control diseases, including distribution of medication, vaccination plans, and allocation of resources (Mostashari and Hartman, 2003).
In recent years, a number of syndromic surveillance approaches have been proposed. According to a study conducted by the Centers for Disease Control and Prevention (CDC) in 2003 (Buehler et al., 2003), roughly 100 sites throughout the country have implemented and deployed syndromic surveillance systems. These systems, although sharing similar objectives, vary in system architecture, information processing and management techniques, and algorithms for anomaly detection, and have different geographic coverage and disease focuses. We see a critical need for an in-depth review that analyzes and evaluates these existing systems and related outbreak modeling and detection work under a unified framework. Such a study presented in an easily accessible manner will be useful for researchers who are working or have an interest in public health surveillance as a review of the state-of-the-art syndromic surveillance research and practice. It will also provide a much-needed comparative study for public health practitioners and offer concrete insights that could help future syndromic surveillance system development and implementation.

This monograph serves to investigate the surveillance capacity and effectiveness of existing syndromic surveillance systems so as to present a synthesized review of the state of the art in syndromic surveillance research and practice and provide insights and guidelines for future research and system implementation. In comparison with several review articles that were published in this area (Bravata et al., 2004; Lober et al., 2002; Mandl et al., 2004; Yan et al., 2006), this monograph, a significantly extended version of a recent review that we completed and published in a journal article format (Yan et al., 2008), focuses on an in-depth description of technical components of syndromic surveillance systems and frames the related research questions from an IT and informatics perspective.

More specifically, this monograph serves the following purposes: (1) to provide an updated review of existing system development efforts and emerging syndromic surveillance techniques; (2) to identify the emerging needs and challenges; (3) to present in a synthesized manner the research and development efforts of public health agencies, research institutions, and the industry from an IT perspective; and (4) to serve as a tutorial for IT researchers interested in the emerging field of syndromic surveillance and infectious disease informatics. This survey aims to help answer the following questions:

- Is syndromic surveillance an effective approach to the public health surveillance problem? To what extent are existing systems already serving the purpose of early event detection, situation awareness, and response facilitation? How can their usability and effectiveness be validated?
• What information sharing, outbreak detection and information access and visualization techniques have been implemented and how do these techniques perform? Are there any technical barriers to the design and implementation of these approaches in public health?

• What is the deployment status of existing syndromic surveillance systems in the United States and other parts of the world? Are there any legal or administrative challenges hindering their wide adoption?

This book investigates a number of public health syndromic surveillance systems and related outbreak modeling and detection research, with the specific emphasis on the most promising practices in applying advanced information technologies to public health surveillance. It is mainly focused on major efforts from the public health agencies, research institutions, and the industry in the United States. Some other countries with major syndromic surveillance practices, including Canada, the UK, Australia, Japan, and Korea, are also included in the survey.

To prepare this book, we have reviewed about 250 publications from 1997 to 2008. To identify related work, we searched archival journals including but not limited to Journal of Biomedical Informatics, Journal of American Medical Informatics Association, Journal of Advances in Disease Surveillance, Journal of Urban Health, Artificial Intelligence in Medicine, and Annual Review of Information Science and Technology. These journal articles were mainly retrieved from online bibliographical databases including PubMed Medline, ScienceDirect, and SpringerLink. Our literature search used both general keywords such as “syndromic surveillance” and “biosurveillance,” and keywords pertaining to various technical aspects of syndromic surveillance such as “outbreak detection,” “spatial surveillance”, and “bioterrorism preparedness.” In addition, we investigated other research outlets, including proceedings and presentation material from various workshops (e.g., Arizona BioSurveillance Workshops 2006, 2007, and 2008, and Rutgers DIMACS Working Group on BioSurveillance Data Monitoring and Information Exchange). User manuals and system brochures that are available electronically (e.g., from state/national health department Web sites) were also studied.

Our work reported in this book aims to be comprehensive and is based on a systematic study of over fifty syndromic surveillance systems. (Our review does not count implementations of one system in multiple sites.) We believe these surveyed systems represent most of the known syndromic surveillance systems for which technical descriptions in varying degrees of detail are available from public sources. Technical approaches or solutions from each system are carefully catalogued and analyzed based on their purpose, input assumed, and output produced. The similarities and differences between these approaches are identified and their relative strengths and
weaknesses summarized. In addition, an attempt has been made to perform a “post analysis,” cutting across all these systems with the objective of assessing the extent to which a particular technical approach has been used to meet a specific functional requirement of syndromic surveillance.

Our discussion of public health syndromic surveillance systems is based on a conceptual framework (Figure 1-1) that views syndromic surveillance as composed of three main functional areas: data sources and collection strategies; data analysis and outbreak detection; and data visualization, information dissemination, and reporting. Most modern syndromic surveillance systems can be conceptualized following this framework.

The first area is primarily concerned with where and how to collect data. The related issues include data entry approaches, data sharing protocols, and transmission techniques. The second area involves modeling, analysis, and data mining approaches to monitor for data anomalies and to discover whether the aberrant data condition is caused by a real change in disease occurrence. The syndrome classification process, a critical step that occurs between data collection and anomaly detection, focuses on classifying the raw, observational data into syndrome groups to provide a meaningful representation with the appropriate level of abstraction and granularity to detect aberrations in any monitored illness. The third area involves data visualization, user interface, and information dissemination functionalities. Public health officials, epidemiologists, and when appropriate, emergency response and homeland security personnel, interact with the syndromic surveillance systems through these components to access detailed information for further investigation, gain situational awareness, make decisions about alert generation and dissemination, and collect information needed for response planning and event management.

Figure 1-1. Conceptual syndromic surveillance system architecture.
This monograph consists of two main groups of chapters. The first group, Chapters 2–6, follows the above framework, discussing various components of syndromic surveillance systems and approaches. The second group, Chapters 7–14, presents integrative case studies based on representative systems and typical application scenarios.

We conclude this introductory chapter by summarizing the key features of each ensuing chapter. In Chapter 2, a summary of syndromic surveillance systems surveyed in our study, most of which have been adopted in real-world applications, is presented. Chapters 3–5 discuss technical material related to data collection, data analysis and outbreak detection, and data visualization and information dissemination, respectively. System assessment and other policy considerations are reviewed in Chapter 6.

From Chapter 7 to Chapter 14, in each chapter, we report a case study with a particular syndromic surveillance system, covering BioSense, RODS, BioPortal, ESSENCE, NYC SS, EARS, Argus, and HealthMap. Chapter 15 concludes this book by discussing critical issues and challenges to syndromic surveillance research and system development, and proposing some future directions.