

Application of Geographic Information Systems to Syndromic Surveillance

Hui Li, Fazlay Faruque, Worth Williams, Richard Finley

Abstract

We have developed a Web-based real-time syndromic surveillance system with geographic information system disease mapping capabilities. The system includes four major components: real-time data collection, syndrome classification, dynamic spatial mapping, and query capabilities. An electronic medical record system at an urban teaching hospital is the source of real-time ICD-9 discharge data which are then mapped into syndrome categories. The GIS disease mapping is a Web-based tool of value to epidemiologists and public health officials for the interpretation and analysis of both routine and outbreak-related health data.

Introduction

Syndromic surveillance is the utilization of crude health data for the rapid detection of disease outbreaks or bioterrorism attacks. Syndromic surveillance systems have been deployed using routinely collected laboratory, pharmacy, or clinical data such as a patient's chief complaint on arrival at an emergency department. Existing systems have focused on data collection methods, characteristics of the data collected, and analytical methods to detect disease outbreaks [1]. Beyond actually detecting an epidemic, the availability of additional demographic information such as age, sex and location is

valuable for understanding the etiology and dynamics of an outbreak. Despite the advantage of electronic data collection and reporting, many local health departments do not currently have adequate resources to manage, analyze, and interpret such large data sets [1]. Geographic Information Systems (GIS) can play an important role in surveillance systems and help decision-makers interpret and analyze these data. In particular, geographic information about the location of cases and their temporal evolution would be invaluable to those responsible for identifying and controlling an outbreak. The use of these systems for routine health and disease surveillance may be even more useful than monitoring for disease outbreaks or bioterrorism attacks.

Our research focuses on implementing Web-based GIS functions into a real-time syndromic surveillance system. The system includes four major components: real-time data collection, syndrome classification, dynamic spatial mapping, and query capabilities. The query functions will allow filtering of the data by syndrome and demographic variables such as age and sex, focus on particular geographic areas including zip code and county, and analysis of temporal trends using user-determined dates. The Web-based nature of the system will allow anyone with access privileges to query the system for epidemiological analysis.

Implementation

Our GIS Web-based surveillance system (GeoMedStat) will (1) map the spatial distribution of infectious diseases of interest in any give time period and; (2) query disease related information in spatial data layers. ArcIMS is a software tool for delivering dynamic maps and GIS data and services via the Web. For rapid application

development, ArcIMS includes Designer, a component to create ArcIMS viewers using pre-built templates. ArcIMS pre-built viewers support only a limited customization capability by setting predefined parameter values. To meet dynamic map requirements, a customized view was developed using JavaScript, HTML, and DHTML to support flexible mapping functions in this system.

1. Data Source

Health care encounter data, in particular emergency department (ED) data, are readily available and are well suited to syndromic surveillance [2]. ED discharge diagnosis data, from May 2000 to February 2005 (a total of 87350 records) obtained from the University of Mississippi Medical Center (UMMC) Emergency Department patient database, were collected and used in this study. The ED data are acquired using an electronic medical record in real time, and the discharge diagnoses (ICD-9) are coded automatically at the time the patient leaves the ED. This obviates the need for the use of the much less specific chief complaint data which are used in many other syndromic surveillance applications. Data elements imported from the ED database include encounter date, patient zip code, city, sex, age, ICD9 code, ICD9 code description, and ICD-9 rank. Patient identification information is excluded. The ICD-9 is an abbreviation of International Classification of Diseases, Ninth Revision, Clinical Modification, which was developed to allow the assignment of codes to diagnoses and procedures associated with hospital utilization in the United States. The ICD-9 codes represent the ED patients' final diagnoses. One patient visit may have multiple diagnoses and associated ICD-9 codes, and each ICD-9 code in a visit was ranked in order, with the primary diagnosis

first and additional diagnoses listed in rank order. The ICD-9 codes were mapped to different syndromes based on ICD9-codes. This will be discussed in more detail. The data could be imported from the ED database into the surveillance application at arbitrary time intervals, but for practical reasons, in the current implementation a time interval of 24 hours was chosen based on ED census data; shorter intervals were not felt to offer any significant advantage.

2. Syndrome classification

Architects of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) system developed a mapping of ICD-9 codes to syndrome categories which has been widely distributed. The Centers for Disease Control and Prevention (CDC) identified eleven syndrome categories and corresponding ICD-9 codes that can be used in syndromic surveillance programs [3]. Syndromic surveillance systems have used ED chief complaints or ED discharge diagnosis data for syndrome categorization. Free-text chief complaints can be grouped into syndromes using a statistical model such as CoCo Bayesian Classifier [4]. However, previous studies have shown that ICD-9 codes more accurately classify patients into syndromes than chief complaints [5]. The ICD-9 codes are often not available in a timely manner however, and therefore many systems use chief complaint data. One of the advantages of our system is the rapid availability of specific ICD-9 discharge diagnoses

Seven syndrome categories (Gastrointestinal, Botulism-like, Hemorrhagic, Respiratory, Neurological, Rash, and Constitutional) were selected for monitoring in our system, but this can be easily modified for any combination of ICD-9 categories. The ED discharge

diagnosis data were automatically grouped into the above seven syndromes based on their corresponding ICD-9 codes. Two kinds of redundancy in the ED data needed to be considered when these data are classified in the syndromic surveillance system. One is that multiple ICD-9 codes for a single patient visit may belong to the same syndrome group. This redundancy will overestimate the number of cases in a syndrome category and is removed programmatically to eliminate multiple insertions (18.9% of the records had redundant ICD-9 codes to syndrome mappings eliminated). Secondly, multiple ICD-9 codes for a visit may belong to different syndrome groups because a patient's symptoms may fall into more than one syndrome. This redundancy includes important information and is left intact in the database.

3. Spatial distribution mapping

Disease mapping is a major focus in spatial epidemiology and is used to provide insight into possible causes of diseases, clusters of diseases across a geographical area, and the evolution of disease outbreaks. A goal of spatial mapping in this study is to identify regions with unusually high numbers of cases during disease outbreaks or a bioterrorism attack (also referred to as "cluster detection"). A second goal is to determine high-risk areas for a disease of interest. This could then be related to other factors such as environmental pollution, weather conditions or demographic factors. The visits for each syndrome were used to map syndrome distribution at the zip code level. Since diseases have different spatial patterns over time, it is important that the system support disease mapping in any given time frame.

4. Query functions

To help users interpret disease patterns, query capabilities are supported in the system. Users can select different data layers to query disease-related information by clicking on a geographical area of interest. Javascript is used to collect query parameters and transfer them into ArcXML format that is then sent to the ArcIMS server for further processing through .NET link using ASP.NET.

The GeoMedStat user interface can be seen from Figure 1.

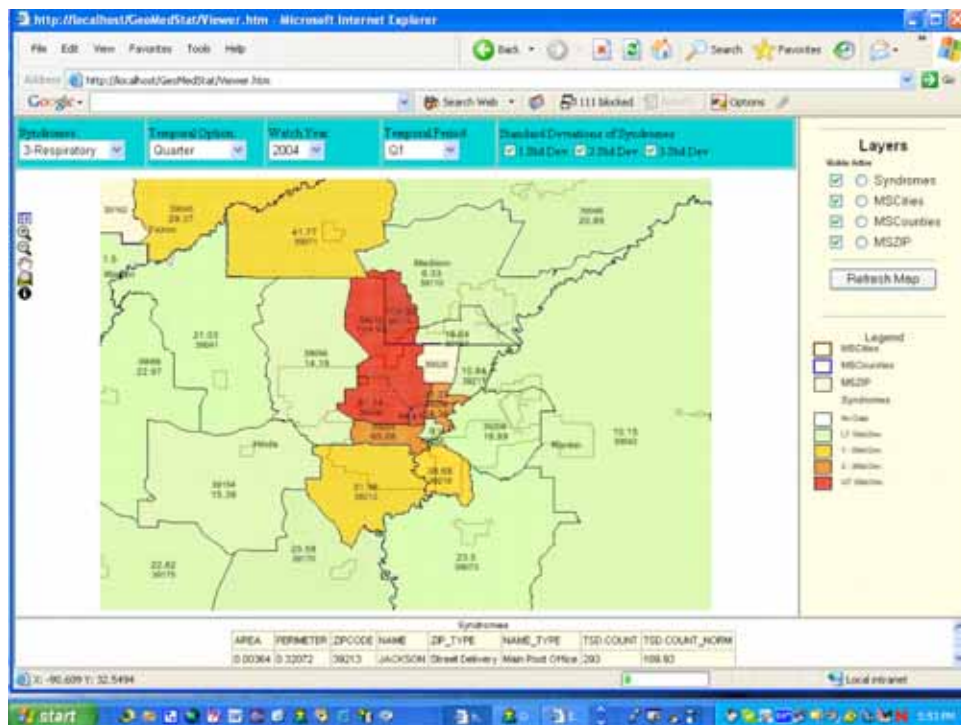


Figure 1. GeoMedStat User Interface

Discussion

Syndromic surveillance is a rapidly evolving discipline driven by timely concerns about emerging disease outbreaks or bioterrorism attacks. Most of the emphasis in the published literature has been on the early detection of outbreaks, and this is certainly a worthwhile goal. Possibly even more important is the use of the associated monitoring tools to analyze an outbreak that has already been detected. This would be an invaluable epidemiological tool to aid in understanding and controlling such an outbreak. Health officials and epidemiologists must consider such problems as dynamics of spread, associated ecological or climatic factors, possible quarantine decisions and resource allocation. For this reason, a real-time geographical picture of the situation is essential. A more prosaic, but probably ultimately more useful, task for such a real-time GIS tool would be for routine epidemiological studies. If such a tool were Web-based and widely available to epidemiologists with differing interests, it would certainly augment public health analysis immeasurably, and point to unsuspected problems much more rapidly than currently possible.

The development of such a tool is beset with several difficult but not insurmountable difficulties. The first is data acquisition. In the past, epidemiological studies have relied on classical methods of physician and hospital reporting, chart reviews, and patient interviews. This is an expensive and time-consuming method that causes it to be very limited in time and scope, and the final release of the data is often delayed by months or years. This is unacceptable when dealing with rapidly evolving situations such as emerging infectious disease epidemics like SARS or a bioterrorism attack with smallpox.

Electronic hospital administrative records have allowed the rapid retrieval of some simple information such as presenting chief complaint and demographic information that can be used for analysis of real-time disease trends. These data are usually collected by non-medical personnel and therefore must be interpreted by algorithms to try to decide what kind of disease it relates to. The gradual deployment of complete electronic medical records has improved the information available with timely and specific diagnosis information. We have the advantage of working in a hospital where a completely electronic medical record has been in place in the emergency department for the past four years. These data are available for importing into our syndromic surveillance system on a real-time basis and include the relevant clinical and demographic information enabling mapping to appropriate syndromes and to our GeoMedStat geographic disease mapping program.

Using GeoMedStat we can now see real-time syndrome data mapped to zip code level within the state over a Web-based interface. Our current research is focused on the best methods for automating the presentation and interpretation of this data. A major problem with the interpretation of spatial mapping is data normalization. There is a large amount of both temporal and spatial variability that must be taken into account. For example, a known temporal variability is the seasonal variation in respiratory diseases with increases during the winter months. Spatial variability is even more problematic. Our hospital is centrally located in the state and draws patients from the entire state. However, the number of patients seen and their severity of illness are associated with the distance the patient must travel to reach our hospital. Rural areas have large variations in population

density that must be considered. Figure 2 shows an example of the temporal and spatial variation of the data for respiratory syndrome over the years 2001 through 2004.

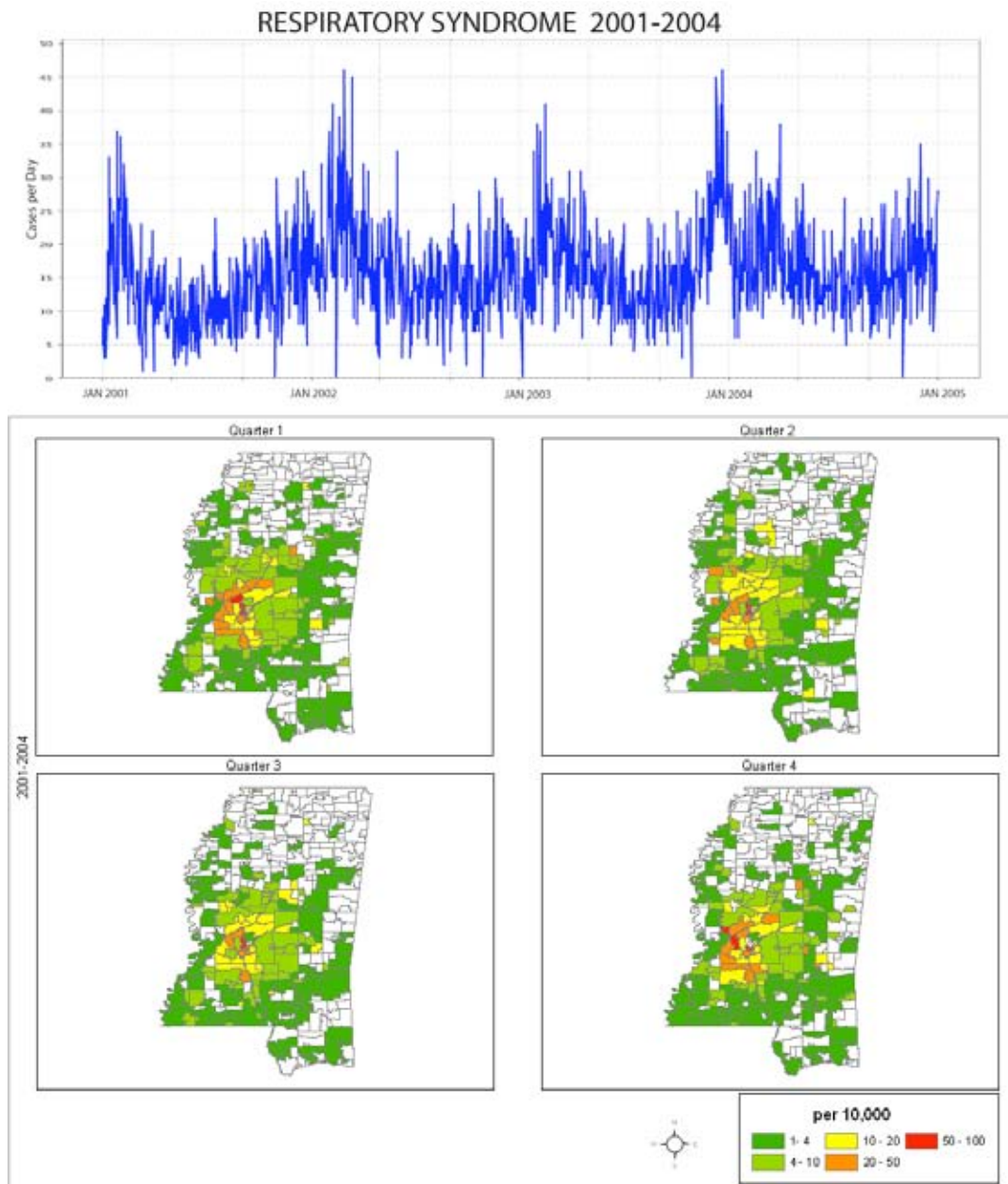


Figure 2. Temporal and spatial variation of the data for respiratory syndrome over the years 2001 through 2004.

These normalization issues are a complex topic, and implementation of algorithms to study the advantages and efficiency of different techniques are currently under active research using tools such as time series analysis, cluster analysis including spatial scan statistics, neural networks, and simulation modeling.

Conclusions

GIS can play an important role in disease outbreak surveillance systems and help decision makers interpret and analyze both routine and outbreak-related health data.

References

1. Bravata, DM, et al. Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Annals of Internal Medicine* 2004; 140(11):910-922.
2. Mandl, KD, et al. Implementing syndromic surveillance: A practical guide formed by the early experience. *Journal of the American Medical Informatics Association* 2004; 11:141-150.
3. Syndrome definitions for diseases associated with critical bioterrorism-associated Agents. Available at: <http://www.bt.cdc.gov/surveillance/syndromedef/index.asp> access.
4. Ivanov, M, et al. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. AMIA Fall Symp. *JAMIA* 2002; (Supplement):345-349.
5. Beitel AJ, Olson KL, Reis BY, Mandl KD. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 2004; 20(6):355-360.
6. Espino, JU, et al., 2004. The RODS open source project: Removing a barrier to syndromic surveillance. *CDC's Morbidity and Mortality Weekly Report* 2004; 53(Supplement 1):32-39

Author information

Hui Li, Ph.D, Department of GIS, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216

Fazary Faruque, Ph.D, Associate Professor, Department of GIS, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216

Worth Williams, Senior Analyst/Programmer, Department of GIS, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216

Richard Finley, M.D, Professor, School of Medicine, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216

