Chapter 3

USING EMERGENCY DEPARTMENT DATA FOR BIOSURVEILLANCE: THE NORTH CAROLINA EXPERIENCE

ANNA E. WALLER*, MATTHEW SCHOLER, AMY I. ISING, and DEBBIE A. TRAVERS

CHAPTER OVERVIEW

Biosurveillance is an emerging field that provides early detection of disease outbreaks by collecting and interpreting data on a variety of public health threats. The public health system and medical care community in the United States have wrestled with developing new and more accurate methods for earlier detection of threats to the health of the public. The benefits and challenges of using Emergency Department data for surveillance are described in this chapter through examples from one biosurveillance system, the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT). ED data are a proven tool for biosurveillance, and the ED data in NC DETECT have proved to be effective for a variety of public health uses, including surveillance, monitoring and investigation. A distinctive feature of ED data for surveillance is their timeliness. With electronic health information systems, these data are available in near real-time, making them particularly useful for surveillance and situational awareness in rapidly developing public health outbreaks or disasters. Challenges to using ED data for biosurveillance include the reliance on free text data (often in chief complaints). Problems with textual data are addressed in a variety of ways, including preprocessing data to clean the text entries and address negation.

* *Carolina Center for Health Informatics, Department of Emergency Medicine. University of North Carolina at Chapel Hill, 100 Market Street, Chapel Hill, NC 27516, USA, awaller@med.unc.edu*

D. Zeng et al. (eds.), *Infectious Disease Informatics and Biosurveillance*,
Integrated Series in Information Systems 27, DOI 10.1007/978-1-4419-6892-0_3,

The use of ED data for public health surveillance can significantly increase the speed of detecting, monitoring and investigating public health events. Biosurveillance systems that are incorporated into hospital and public health practitioner daily work flows are more effective and easily used during a public health emergency. The flexibility of a system such as NC DETECT helps it meet this level of functionality.

1. INTRODUCTION

Biosurveillance is an emerging field that provides early detection of disease outbreaks by collecting and interpreting data on a variety of public health threats, including emerging infectious diseases (e.g., avian influenza), vaccine preventable diseases (e.g., pertussis) and bioterrorism (e.g., anthrax). With the Centers for Disease Control and Prevention's (CDC) initial focus on bioterrorism preparedness at the state and local level in 1999 and the subsequent anthrax outbreak of 2001, the public health system and medical care community in the United States have wrestled with developing new and more accurate methods for earlier detection of threats to the health of the public. Earlier detection, both intuitively and as illustrated through predictive mathematical models, is believed to save lives, prevent morbidity and preserve resources (Kaufman et al., 1997). Biosurveillance systems use health-related data that generally precede diagnoses and that signal a sufficient probability of a case or an outbreak to warrant further public health response (Buehler et al., 2004).

Rapid detection of disease outbreaks rests on a foundation of accurate classification of patient symptoms early in the course of their illness. Electronic emergency department (ED) records are a major source of data for biosurveillance systems because these data are timely, population-based and widely available in electronic form (Lober et al., 2002; Teich et al., 2002). There are more than 115 million ED visits annually in the United States, and EDs represent the only universally accessible source of outpatient healthcare that is available 24 h a day, 7 days a week (Nawar et al., 2007). EDs see patients from all age groups and socioeconomic classes. Patients may present with early, nonspecific symptoms or with advanced disease. The accessibility of EDs provides a likely healthcare setting for many of the patients involved in a disease outbreak of public health significance. In recent years, EDs have steadily adopted electronic medical records technology (Hirshon, 2000), which has facilitated the replacement of drop-in manual surveillance using ED data with ongoing, real-time surveillance. ED data have been shown to detect outbreaks 1–2 weeks earlier than traditional public health reporting channels (Heffernan et al., 2004; Lober et al., 2002; Tsui et al., 2001; Wagner et al., 2004).

The ED data elements that are used for biosurveillance include the chief complaint (a brief description of the patient's primary symptom(s)), the triage nurse's note (an expansion of the chief complaint that includes the history of present illness), other clinical notes (e.g., physician and nurses' progress and summary notes), initial measured temperature, and diagnosis codes. The most widely used ED data element is the chief complaint because it is recorded electronically by most EDs and may precede entry of a diagnosis or transcription of physician notes by days or weeks (Travers et al., 2003, 2006). The triage note increases the amount of data available, which makes it more likely that biosurveillance algorithms will detect disease outbreaks. Triage notes are becoming more available in electronic form, and one study found that adding triage notes increased the sensitivity of outbreak detection (Ising et al., 2006).

Several challenges to using ED data for biosurveillance have been identified (Hirshon, 2000; Varney & Hirshon, 2006), including costs to EDs and public health, the lack of standardization of ED data, and security and confidentiality. Many EDs still document patient symptoms manually; even when the data are electronic, they are often entered in free text form instead of using standardized terms. Timeliness is also a concern; while some ED data elements are entered into electronic systems at the start of the ED visit, other elements are added hours, days or even weeks later. Even though there is no formal standard or best practices dictating how soon data should be available after an ED visit or other health system encounter for early detection, most surveillance systems aim for near real-time data, available within hours.

The benefits and challenges of using ED data for surveillance will be described in more detail through examples from one biosurveillance system, the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT). NC DETECT evolved from a pilot project in 1999 to demonstrate the collection of timely, standardized ED data for public health surveillance and research. NC DETECT has since grown to incorporate ED visit data from 98% of 24/7 acute care hospital EDs in the state of North Carolina and has developed and implemented many innovative surveillance tools, including the Emergency Medicine Text Processor (EMT-P) for ED chief complaint data and research-based syndrome definitions. NC DETECT now provides twice-daily ED data feeds to CDC's BioSense and has over 200 registered users at the state, regional and local levels across North Carolina. This chapter will review the use of ED data for biosurveillance, including appropriate case studies from NC DETECT.

2. LITERATURE REVIEW/OVERVIEW OF THE FIELD

2.1 History of ED Data Use for Biosurveillance

ED data have been collected for decades for a variety of public health surveillance needs and have been incorporated into electronic systems designed to analyze data related to trauma, injury and substance abuse, among others. Public health officials have used event-based or drop-in biosurveillance systems that include ED data during major events, including the Olympic Games, political conventions, heat waves, after major hurricanes, and after the identification of known widespread outbreaks (Weiss et al., 1988; Davis et al., 1993; Lee et al., 1993; Rydman et al., 1999). Many of these systems have required users to do manual abstractions from medical charts or to enter data into stand-alone systems for specific symptoms of interest. For example, the EMERGEncy ID NET program, established in the late 1990s, created a network of select EDs to manually collect data to study syndromes related to emerging infections of interest to the CDC, using paper forms and standardized computer screens (Talan et al., 1998).

Secondary data, data that are generated as part of normal patient treatment and billing, are generally extracted from hospital information system(s) through automated programs either in real-time (at the time of record generation) or near real-time (hourly, every 12 h, daily). Surveillance systems that use secondary data are intended to be less burdensome to ED staff and less costly than systems requiring manual abstraction (Rodewald et al., 1992). This methodology of ED data collection has become standard practice for biosurveillance systems using ED data, including NC DETECT, RODS, ESSENCE and EARS, among others (Hutwagner et al., 2003; Ising et al., 2006; Lombardo, 2003; Wagner et al., 2004, Waller et al., 2007). While there are several different approaches to automated extraction programs, most rely either on delimited text batch files or HL7 messages.

2.2 Current Status of ED Data Use for Biosurveillance

According to the International Society for Disease Surveillance (ISDS) State Syndromic Surveillance Use Survey, 76% of responding states ($n = 33$) performing syndromic surveillance use ED data (http://isds.wikispaces.com/Registry_Project, accessed June 4, 2008). (Figure 3-1).

While most states and regions rely on ED chief complaint data, there is interest in increasing the number of ED data elements collected, as evidenced by the American Health Information Community's Biosurveillance Minimum Data Set (http://www.hhs.gov/healthit/ahic/materials/meeting10/

bio/BDSG_Minimum_DataSet.doc, accessed June 4, 2008). The recommendations from the American Health Information Community include additional emergency department data elements, such as triage notes, vital signs and ICD-9-CM-based diagnoses. The Biosurveillance Minimum Data Set is currently under formal evaluation for its utility at CDC-funded sites in Indiana, New York and Washington/Idaho.

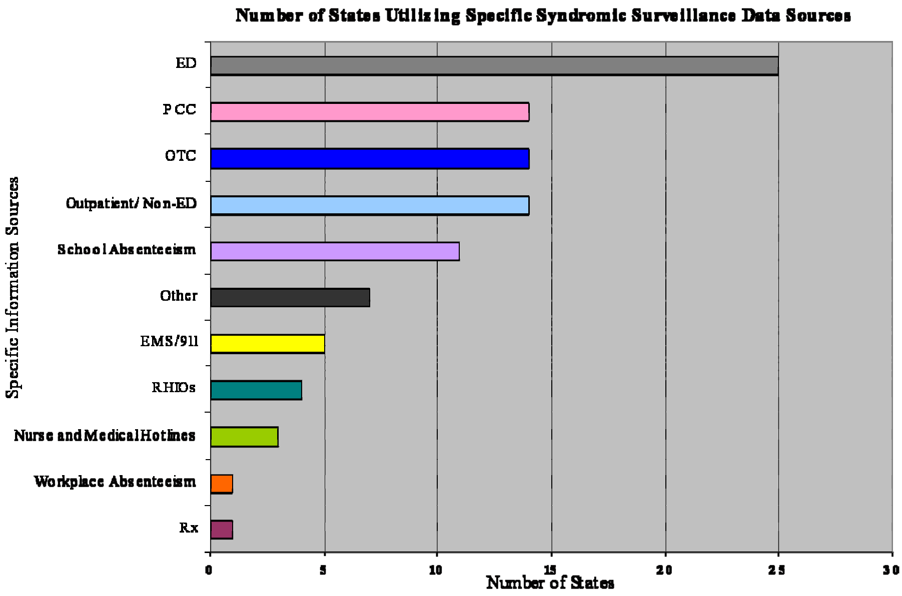

Figure 3-1. ISDS state syndromic surveillance use survey: 41 respondents; 33 use syndromic surveillance

2.3 Infectious Disease Syndrome-Based Surveillance Using ED Data

While infectious disease surveillance has traditionally relied on laboratory results and the reporting of mandated reportable conditions by medical practitioners, ED visit data and timely symptom-based analysis provide additional means for early identification of infectious disease outbreaks. Areas of particular interest include the CDC's list of potential bioterrorism agents, as well as post-disaster (e.g., hurricane, earthquake, chemical spill) surveillance (CDC, 2003). The ability to create effective syndromes to use with ED visit data is of paramount importance to their timely use for public health surveillance.

The structure of syndrome definitions used in biosurveillance is dependent on the design of the system and the nature of the data under surveillance. Individual systems use different methods to identify specific disease symptoms in the chief complaint and triage note data. This includes deterministic methods, such as keyword searching, and probabilistic methods, such as naïve Bayesian and probabilistic machine learning (Espino et al., 2007). Syndrome definitions then classify records into syndromic categories based on which symptoms are identified. To date, no best practices exist to guide syndrome definition development and evaluation (Sosin & DeThomasis, 2004). Which syndromes are monitored and which symptoms are associated with each syndrome varies according to the system under consideration. Furthermore, syndrome structure may vary depending upon which data elements, in addition to chief complaint, are available and their timeliness. Syndrome structure refers to how many symptoms are required, within which data fields they must be found, and which Boolean operators are employed to determine whether a certain record matches a particular syndrome.

2.4 ISDS Consultative Syndrome Group

In September 2007, the ISDS sponsored a consultative meeting on chief complaint classifiers and standardized syndromic definitions (Chapman & Dowling, 2007). At this meeting, representatives from eleven syndromic surveillance systems throughout the country, including NC DETECT, met to discuss which syndromes they monitor and which chief complaint-based clinical conditions they include in each syndrome. Clinical conditions are medical concepts which may be represented by multiple possible data inputs to the system. For example, the concept of "dyspnea" may be represented by signs and symptoms of dyspnea as recorded in an ED chief complaint or triage note field (e.g., "Shortness of Breath" (SOB)), a clinical observation such as an abnormal vital sign (e.g., low oxygen saturations or increased respiratory rate), clinical findings (e.g., abnormal breath sounds on pulmonary exam), abnormal lab findings (e.g., abnormal ABG or positive culture results), imaging studies (e.g., infiltrate on chest x-ray), or certain ICD-9-CM diagnosis codes (e.g., 486, pneumonia). The meeting participants reached consensus on best practices for which clinical conditions to associate with each of six different syndromes (sensitive and specific versions of respiratory syndrome and gastrointestinal syndrome, constitutional syndrome and influenza-like illness syndrome). Through online collaboration and periodic conference calls, the group continues the process of defining specific chief complaint search terms/keywords which best represent these clinical conditions.

3. TECHNICAL APPROACHES FOR GROUPING ED DATA INTO SYNDROMES FOR BIOSURVEILLANCE

While the process of identifying specific chief complaint search terms/ keywords to group into syndromes presents several technical challenges, the timeliness of chief complaints outweighs the benefits of any standardized data that are not available within hours of the ED visit. Textual data such as chief complaint and triage note present several problems, including misspellings and use of ED-specific and locally-developed acronyms, abbreviations and truncations (Travers & Haas, 2003). There are two main approaches to dealing with the variability in textual surveillance data: (1) incorporating keywords in the actual search query statements; or (2) preprocessing the data. In systems that build various keyword searches (e.g., lexical variants, synonyms, misspellings, acronyms, and abbreviations) into the actual surveillance tools, elaborate search statements are constructed, employing statistical software such as SAS (Cary, NC), or standard query language (SQL, Microsoft, Redmond, WA) (Forbach et al., 2007; Heffernan et al., 2004). In systems with preprocessors, the data are cleaned prior to application of a syndromic classification algorithm (Mikosz et al., 2004; Shapiro, 2004). The preprocessors clean text entries, replacing synonyms and local terms (e.g., *throat pain, throat discomfort, ear/nose/throat problem*), as well as misspellings, abbreviations, and truncated words (e.g., *sorethroat*, *sore throaf*, *soar throat*, *ST*, *S/T*, *sore thrt*, *sofe throat, ENT prob*), with standard terms (e.g., *sore throat*) or standard identifiers (e.g., UMLS® concept unique identifier C0242429) (NLM, 2007). Preprocessors often include normalization tools to eliminate minor differences in case, inflection and word order and to remove stop words (NLM, 2006).

While there is no consensus about which approach is best, many biosurveillance programs are implementing preprocessors to improve operations (Dara et al., 2007; Hripscak et al., 2007; Komatsu et al., 2007). Use of preprocessors can streamline maintenance of existing and development of new surveillance queries. Query processing time is also faster, resulting in better overall biosurveillance system performance. One such preprocessor is the Emergency Medical Text Processor (EMT-P), which was developed to process free text chief complaint data (e.g., *chst pn, ches pai, chert pain, CP, chest/abd pain*) in order to extract standard terms (e.g., *chest pain*) from emergency departments (Travers & Haas, 2003). EMT-P has been evaluated by biosurveillance researchers in Pennsylvania and found to improve syndromic classification (Dara et al., 2007). The developers continue to improve EMT-P and have made it publicly available (Travers, 2006).

3.1 Dealing with Negation

While clinical text such as triage notes can improve the accuracy of keyword-based syndrome queries, the data require processing to address negated terms (Ising et al., 2006; Hripcsak et al., 2007). One study evaluated NegEx, a negation tool developed at the University of Pittsburgh (Chapman et al., 2001). NegEx is a simple regular expression algorithm that filters out negated phrases from clinical text. The NegEx system was modified (to include the negation term (-)) and then combined with selected modules from EMT-P that replaced synonyms (e.g., *dec loc* with *consciousness decreased*) and misspellings (*nasaue* with *nausea*) for use in NC DETECT. The pilot results show that this combination of EMT-P and NegEx leads to more accurate negation processing (Ising et al., 2007).

3.2 Issues with Diagnosis Code Data

Another ED data element available for biosurveillance is the final diagnosis, which is widely available in electronic form and is standardized using the International Classifications of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) (USDHHS, 2006). All EDs generate electronic ICD-9-CM diagnoses as they are required for billing purposes (NCIPC, 1997; Travers et al., 2003). There is, however, some evidence that diagnosis data are not always available in a timely manner. In contrast to chief complaint data, which are generally entered into ED information systems by clinicians in real-time, ICD-9-CM diagnoses are often entered into the system by coders well after the ED visit. Sources of ICD-9-CM data may vary, which may influence the quality of the data. Traditionally, diagnoses have been assigned to ED visits by trained coders who are employed by the hospital and/or physician professional group. The primary purpose of the coding is billing, as opposed to secondary uses such as surveillance. Recently, emergency department information systems (EDIS) have come on the market that allow for diagnosis entry by clinicians. These systems typically include drop-down boxes with text that corresponds to ICD-9-CM codes; clinicians can then select a "clinical impression" at the end of the ED visit and the corresponding ICD-9-CM code becomes part of the EDIS data available for surveillance.

In a 2003 study of regional surveillance systems in North Carolina and Washington, biosurveillance developers found that over half of the EDs did not have electronic diagnosis data until 1 week or more after the ED visit (Travers et al., 2003). In a follow up study, researchers prospectively measured the time of availability of electronic ICD-9-CM codes in NC DETECT for all ED visits on 12/1/05 (Travers et al., 2006). The study confirmed that fewer than half of the EDs sent diagnoses within 1 week of the visit, and that it took 3 weeks to get at least one diagnosis for two-thirds of the visits. Seven (12%) of the hospitals had diagnoses for less than two-thirds of their ED visits at the 12 week mark. Diagnosis data are universally available from NC EDs, and studies have shown that ICD-9-CM data alone or in combination with chief complaint data are more valid than chief complaint data alone for syndromic surveillance (Beitel et al., 2004; Fleischauer et al., 2004; Reis & Mandl, 2004). This study corroborated the earlier study, however, that indicated the majority of North Carolina hospitals cannot send diagnosis data soon enough for timely, population-based biosurveillance.

4. BIOSURVEILLANCE IN NORTH CAROLINA

The North Carolina Emergency Department Database (NCEDD) project, spearheaded by the UNC Department of Emergency Medicine (UNC DEM) in 1999, laid the groundwork for electronic ED data collection in North Carolina by developing best practices for collecting and standardizing quality ED data. The focus on using ED data in North Carolina specifically for biosurveillance began in 2002 through a collaboration between UNC DEM and the North Carolina Division of Public Health (NC DPH). In 2004, a partnership between the North Carolina Hospital Association (NCHA) and NC DPH was instrumental in establishing ED data transmissions from the hospitals not yet participating in NC DETECT, including support for a new law mandating reporting as of January 1, 2005 (North Carolina General Statute 130A, http://www.ncleg.net/EnactedLegislation/Statutes/HTML/ ByChapter/ Chapter_130A.html, accessed January 17, 2008). As of October 1, 2010, there are 112/114 (98%) acute care, 24/7 hospital EDs submitting over 11,000 new visits to NC DETECT daily. These data are also transmitted twice daily to CDC's BioSense program.

In addition to ED data, NC DETECT receives data hourly from the statewide Carolinas Poison Center, and daily data feeds from the statewide Emergency Medical System (EMS) data collection center, a regional wildlife center, selected urgent care centers, and three laboratories of the NC State College of Veterinary Medicine (microbiology, immunology and vector-borne diseases laboratories) (Waller et al., 2008).

NC DETECT assists local, regional and state public health professionals and hospital users in identifying, monitoring, and responding to potential terrorism events, man-made and natural disasters, human and animal disease outbreaks and other events of public health significance. This system makes it possible for public health officials to conduct daily surveillance for clinical syndromes that may be caused by infectious, chemical or environmental agents. Suspicious syndromic patterns are detected using the CDC's EARS CUSUM algorithms, which are embedded in the NC DETECT Java-based Web application. The system also provides broader surveillance reports for ED visits related to hurricanes, injuries, asthma, vaccine-preventable diseases and environmental health (Waller et al., 2008). Role-based access provides hospital and public health access to NC DETECT data at local, regional and state levels; multi-tiered access provides tight controls on the data and allows all types of users to access the system, from those who need only an aggregated view of the data, to those who are able to decrypt sensitive protected health information when needed for investigation.

NC DETECT provides an excellent example of an early event detection and situational awareness system using ED visit data for disease surveillance. It is well established, statewide, and utilized daily by a variety of public health practitioners. A recently completed study found that NC ED information in NC DETECT compared favorably with national estimates of ED data made by the National Hospital Ambulatory Care Survey, despite differences in data collection methods (Hakenewerth et al., 2008). This finding is an indication of a well designed and robust system (Aylin et al., 2007).

4.1 History of Syndrome Definitions in NC

The syndromes monitored in NC DETECT are derived from the CDC's text-based clinical case definitions for bioterrorism syndromes (CDC, 2003). These syndromes were selected because they encompass both potential bioterrorism-related and community acquired disease processes. They include botulism-like illness (botulism), fever-rash (anthrax, bubonic plague, smallpox, tularemia, varicella), gastrointestinal (gastrointestinal anthrax, food/water-borne gastrointestinal illness, viral gastroenteritis), influenza-like-illness (epidemic influenza, pandemic influenza), meningoencephalitis (meningitis, encephalitis) and respiratory (respiratory anthrax, pneumonic plague, tularemia, influenza, SARS). Clinical case definitions are converted to syndrome definitions by expressing them in SQL, in most cases requiring both a syndrome specific and a constitutional keyword in either the chief complaint or triage note field. For example, a record containing the syndrome specific term "cough" and the constitutional term "fever" would match the respiratory syndrome. Documentation of a fever by vital sign measurement in the ED is also accepted in lieu of a constitutional keyword. The SQL code is written to identify common synonyms, acronyms, abbreviation, truncations, misspellings and negation in the free text data. The NC DETECT syndrome definitions have been modified over several years in an iterative fashion according to the knowledge and experience of the NC DETECT Syndrome Definition Workgroup. This workgroup meets monthly and includes public health epidemiologists who are regular users of NC DETECT for biosurveillance at the state and local levels, as well as clinicians and technical staff at NC DETECT. The continued improvement of the syndrome definitions for the purposes of syndromic surveillance requires more than this local effort, however. It requires collaboration with other system developers to determine the best practices nationally, as well as evidence-based research to support existing practices and/or develop new methodologies.

4.2 The Importance of Data Quality

The effectiveness of systems such as NC DETECT depends on the quality of the data provided by the data sources and on the system's capacity to collect, aggregate and report information. Perfect data, however, rarely exist and there are no perfect data systems. Thus, assessing and improving data quality must be ongoing tasks.

In NC DETECT, both automated and manual data quality checks are conducted daily and weekly. A Data Quality Workgroup meets monthly to review ongoing data quality concerns and strategize ways to address them. Major data quality issues range from failure to submit data at all to incorrect mapping of county of residence to extended delays in submitting diagnosis and procedure code data. Issues of particular concern include incomplete daily visit data, missing chief complaint data, failure to submit data in a timely fashion, and submission of invalid codes. Successfully addressing ED data quality issues requires constant monitoring of the data and ongoing communication with the hospitals submitting the data to NC DETECT.

4.3 NC DETECT Case Studies

NC DETECT has been used for a variety of public health surveillance needs including, but not limited to, early event detection, public health situational awareness, case finding, contact tracing, injury surveillance and environmental exposures (Waller et al., 2008). Those disease outbreaks that are first identified by traditional means are still aided by ED-based surveillance systems for identifying additional suspected cases and documenting the epidemiology of the affected individuals.

4.3.1 Public Health Surveillance During and After Hurricanes

Several major hurricanes have made landfall or passed through North Carolina in the past 10 years, including Floyd in 1999, Isabel in 2003, and Ophelia in 2005. In addition, hundreds of Katrina evacuees entered North Carolina in August and September 2005. While ED data were used in all instances to monitor the hurricanes' effects, the methodologies used show the evolution of ED data collection for public health surveillance in North Carolina.

In the fall of 1999, Hurricane Floyd, preceded by Hurricane Dennis and followed by Hurricane Irene, caused massive rainfalls that flooded eastern regions of North Carolina along the Neuse, Tar, Roanoke, Lumbar and Cape Fear Rivers. As NCEDD was still in early development in 1999, a disaster response team and ED staff in 20 hospitals worked together to define and apply standardized illness and injury classifications in order to conduct surveillance for the period of September 16 to October 27, 1999 and to compare results to similar periods in 1998. These classifications were applied manually based on diagnosis or chief symptoms for each patient visit abstracted from daily ED logs. Based on these analyses, Hurricane Floyd resulted in increases in insect stings, dermatitis, diarrhea and mental health issues as well as hypothermia, dog bites and asthma. The leading cause of death related to the storm was drowning (CDC, 2000). Surveillance for this time period required the dedicated efforts of EIS officers, medical students and other field staff, as well as ED staff and public health epidemiologists over an extended time period.

NC DPH conducted similar surveillance after Hurricane Isabel in 2003, manually surveying 35 hospitals to document hurricane-related morbidity and mortality (Davis et al., 2007). Officials updated the survey instrument to collect more information on injuries and asked hospitals to complete and fax the information to NC DPH. While less labor intensive overall than the surveillance that took place after Hurricane Floyd, the reliance on ED staff to provide information resulted in a relatively slow and extended collection of data from EDs.

Federal officials evacuated two large groups to North Carolina from Katrina-hit areas of the Gulf Coast in August and September 2005. For this event, NC DPH relied on NC DETECT and hospital-based public health epidemiologists in Wake and Mecklenburg counties for ED data collection. While the epidemiologists at two hospitals were able to identify more Katrina-related visits ($n = 105$) than the automated NC DETECT reports ($n = 90$), the NC DETECT reports required no manual tabulations and took only 2 h to develop and implement. In addition, the epidemiologist count included patients not included in the NC DETECT database, such as patients who were directly admitted to the hospital, without receiving treatment in the ED (Barnett et al., 2007). Furthermore, during the time the Katrina evacuee visits were being monitored, Ophelia approached the NC coast, where it stalled and resulted in the evacuation of coastal communities for several days. NC DETECT was used to monitor Ophelia-related ED visits simultaneously with the Katrina evacuee monitoring effort.

While manual tabulations may result in greater specificity, near real-time automated ED data collection for post-disaster surveillance provides a very low cost approach for monitoring the public's health if a system is already in place and operational. Queries can be continually refined to capture specific keywords in the chief complaint and triage note fields without added burden to hospital and/or public health staff. ED data collection provides an excellent complement to rapid needs assessments and other on-the-ground survey tools. Automated ED data collection assumes that EDs remain operational and that computerized health information systems continue to be used in times of mass disaster, an assumption that has not yet been put to the test in North Carolina.

4.3.2 Influenza

The NC DETECT influenza-like illness (ILI) definition, based on ED data, is used to monitor the influenza season in NC each year. The ED ILI definition follows the same trend as North Carolina's traditional, manually tabulated Sentinel Provider Network but is available in near real-time, as shown in Figure 3-2.

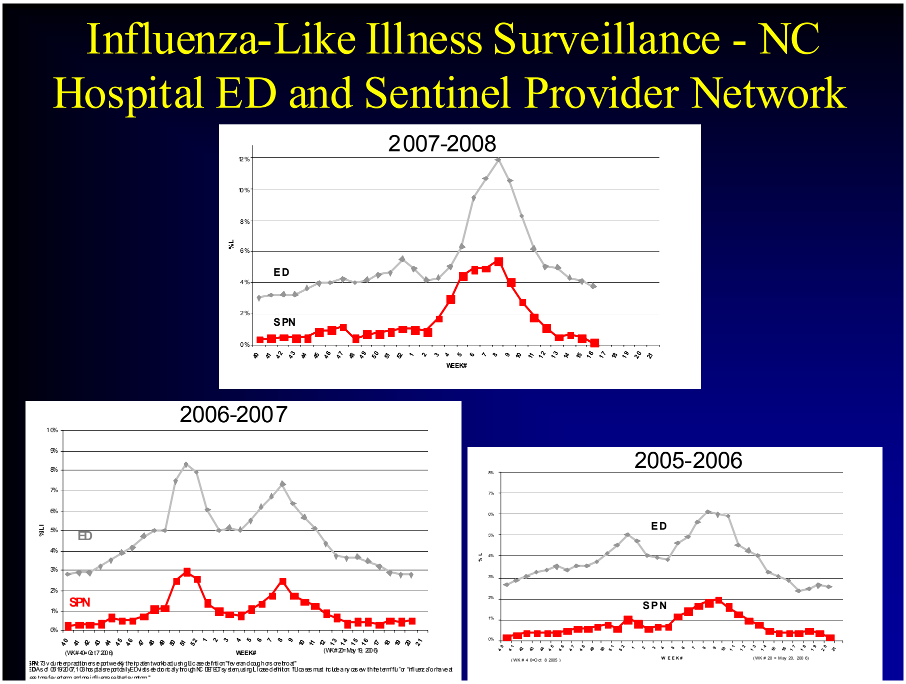

Figure 3-2. Comparing ED ILI (ED) from NC DETECT to the traditional sentinel provider network (SPN)

While North Carolina continues to maintain its Sentinel Provider Network, monitoring influenza with ED data provides several superior surveillance capabilities. In addition to timeliness, collecting ED data for influenza surveillance allows jurisdictions to assess impact on populations rather than samples, test case definition revisions on historical data, stratify ED visits by disposition type (admitted vs. discharged) and incorporate age stratification into analyses. The use of age groups in influenza surveillance has been shown to provide timely and representative information about the age-specific epidemiology of circulating influenza viruses (Olsen et al., 2007). Several states and large metropolitan areas, along with North Carolina, transmit aggregate ED-based ILI counts by age group to an ISDS-sponsored proof-of-concept project called the Distributed Surveillance Taskforce for Real-time Influenza Burden Tracking (DiSTRIBuTE). Although the ILI case definitions are locally defined, the visualizations that DiSTRIBuTE provides show striking similarities in ILI trends across the country (http://www.syndromic.org/projects/DiSTRIBuTE.htm).

4.3.3 Early Event Detection

While syndromic surveillance systems have clearly shown benefit for public health situational awareness and influenza surveillance, early event detection has been more of a challenge. Symptom-based detection systems are often overly sensitive, resulting in many false positives that can drain limited resources (Baer et al., 2007; Heffernan et al., 2004). Hospital and public health users who incorporate syndromic surveillance into their daily work flows, however, are able to accommodate these false positives more efficiently and still derive benefit from monitoring ED data for potential events. Investigating aberrations based on ED data that do not result in detecting an outbreak can still be important to confirm that nothing out of the ordinary is occurring. A recent investigation of gastrointestinal signals in Pitt County, North Carolina, for example, resulted in more active surveillance by the health department (checking non-ED data sources for similar trends) and the hospital (increased stool testing), as well as a health department press release promoting advice for preventing foodborne illnesses. Although a true outbreak or signal causative agent was not detected, this work results in improved coordination and communication among the hospital, healthcare providers and health department, which will make collaboration more efficient in any future large scale response efforts.

4.3.4 Bioterrorism Agent Report

To complement the more sensitive symptom-based syndromes, system developers may also include reports looking for specific mention of Category

A bioterrorism agents, such as anthrax, botulism, etc. In NC DETECT, for example, the Bioterrorism (BT) Agent Case Report searches for keywords and ICD-9-CM diagnoses related to 21 different bioterrorism agent groups. A statewide search on all 21 agents on average returns only ten cases (averaging one case a day over 10 days). In comparison to the specificity of this report, a statewide search on botulism-like illness for 10 days in NC DETECT produces approximately 200 cases while a search on a broad definition of gastrointestinal illness produces approximately 16,000 cases statewide over a 10-day period.

While the BT agent case report does include false positive cases, it provides an effective, unobtrusive monitoring mechanism that complements the astute clinician. It is also an important backup when notifiable diseases go unreported to the public health department, which actually occurred in March 2008 with a single case of tularemia.

4.3.5 Case Finding & Infectious Disease Outbreak Monitoring

Similar to the periods during and after natural disasters, monitoring ED data during a known infectious disease outbreak can assist with case finding and contact tracing. During known outbreaks, NC DETECT is used to identify potential cases that may require follow up. To assist in this effort, the NC DETECT Custom Event Report allows users to request new reports in just 2 h, with specific keyword and/or ICD-9-CM diagnostic criteria (Ising et al., 2007). This report has assisted North Carolina's public health monitoring in several events, including, but not limited to, nationwide recalls of peanut butter (February 2007), select canned foods (July 2007), nutritional supplements (January 2008), as well as localized Hepatitis A (January 2008) and Listeriosis (December 2007) outbreaks. Allowing users to access reports with very specific keywords (e.g., "peanut," "canned chili," "selenium") provides them with an efficient, targeted mechanism for timely surveillance of emerging events, all with the intention of reducing morbidity and mortality.

4.3.6 Infectious Disease Retrospective Analyses

When syndromic surveillance systems collect ICD-9-CM diagnosis codes in addition to chief complaints, users can conduct retrospective analysis effectively. For example, users can search on the ICD-9-CM code V04.5 (need for prophylactic vaccination and inoculation against certain viral diseases: rabies) to review how many ED patients received rabies prophylaxis in a given time period. Using the ED chief complaint, users can go a step further and view how many ED patients with chief complaints related to animal bites/animal exposures were NOT documented as having received a V04.5 code. Investigation of the results may reveal hospital coding errors or hospital practices that are not in line with public health requirements that can then be corrected.

4.3.7 Injury Surveillance

The Injury and Violence Prevention Branch of NC DPH has added ED data from NC DETECT to its data sources for injury surveillance efforts. In addition to ED visit data, they also use hospital discharge, death certificate, and medical examiner data. Injury surveillance efforts involving ED data have included falls, traumatic brain injury, fire-related injury, self-inflicted injury, heat-related visits, and unintentional drug overdoses. Furthermore, they have used ED data when working with trauma Regional Advisory Committees to evaluate injury patterns and are exploring the possibility of incorporating ED data into NC's violent death reporting system. While ED data have long been used for injury surveillance, the availability of near real-time data provides opportunities for more timely documentation of intervention outcomes.

4.4 Conclusions and Discussion

ED data are a proven tool for biosurveillance, and the ED data in NC DETECT have proved to be effective for a variety of public health uses, including surveillance, monitoring and investigation. Biosurveillance systems that are incorporated into hospital and public health practitioner daily work flows are more effective and easily used during a public health emergency. The flexibility of a system such as NC DETECT helps it meet this level of functionality.

4.5 Evaluation of NC DETECT

Any surveillance system should undergo rigorous evaluation to make sure it is meeting user needs effectively and efficiently. The ED data stream of NC DETECT has undergone two such evaluations. In 2007, it was evaluated by the North Carolina Center for Public Health Preparedness at the charge of the NC DPH. The evaluation was designed to determine the usefulness of the ED data and the ease with which it is used for both real-time and non-real-time public health surveillance activities. Interviews were conducted with stakeholders to learn about the specifics of the ED data, data flow, and the aberration detection algorithms. In addition, local, regional and state public health authorities, as well as hospital-based public health epidemiologists (PHEs), were asked to complete a Web-based survey about their experience using the ED data via NC DETECT. Key findings included:

- ED data permit public health authorities to identify human health events as a result of bioterrorism, natural or accidental disaster, or infectious disease outbreak, but the rapidity of detection is contingent on the extent of the event and affected individuals, the ability of chief complaint data to be successfully transmitted to NC DETECT in a timely manner, and the frequency and timing of aberration detection and investigation by public health authorities;
- The NC statute mandating provision of ED visit data for public health surveillance and the availability of UNC DEM staff to provide technical and analytical expertise have been instrumental in assuring that timely, quality data are available for public health surveillance;
- ED data are useful to public health authorities;
- The system showed a low positive predictive value (PPV), indicating that users must examine a large number of false positives in order to identify a potentially true threat to public health.

Based on these findings, this evaluation recommended additional efforts to encourage public health authorities to routinely use the ED data, increased communication among hospitals, business organizations and public health authorities, examination and evaluation of different aberration detection algorithms, and a cost-benefit study of using ED data for public health surveillance.

A second evaluation of the emergency department data stream of NC DETECT was conducted in 2007 by the Research Triangle Institute to assess the impact of this biosurveillance system on public health preparedness, early detection, situational awareness, and response to public health threats. This study used key informant interviews and found the following:

- Biosurveillance has been used in North Carolina for situational awareness and early detection of disease outbreaks;
- Public health epidemiologists in hospitals and regional state-based response teams have integrated use of NC DETECT with traditional surveillance activities;
- Biosurveillance has added timeliness and flexibility to traditional surveillance, increased reportable disease reporting and case finding, and increased public health communication.

This evaluation recommended the addition of urgent care center data to complement the ED visit data for biosurveillance and exploring the use of diagnosis data, when available in a timely manner, to minimize false positive alerts.

5. CONCLUSION

A distinctive feature of ED data for surveillance is their timeliness. With electronic health information systems, these data are available in near real-time, making them particularly useful for surveillance and situational awareness in rapidly developing public health outbreaks or disasters. The use of ED data for public health surveillance can significantly increase the speed of detecting, monitoring and investigating public health events. Combined with other timely data sources such as data from poison centers, EMS, ambulatory care data, and animal health data, ED data analyses are an important source of information for mitigating the effects of infectious disease.

ACKNOWLEDGEMENTS

Financial support for NC DETECT is currently provided by the North Carolina Division of Public Health Contract 01322–11 and the Department of Emergency Medicine of the University of North Carolina at Chapel Hill. The authors acknowledge the collaboration of Lana Deyneka, MD, MPH, of the North Carolina Division of Public Health, and Scott Wetterhall, MD, MPH, and Anne-Lyne McCalla, MPH, of RTI, International in the syndromic surveillance research and development efforts of NC DETECT through BioSense grant R01 PH000038–0.

QUESTIONS FOR DISCUSSION

1. Are timely ED data systems for public health surveillance cost effective? How would you measure this?
2. How can biosurveillance systems and electronic lab reporting for reportable conditions best complement each other?
3. What other data sources could and should be used with ED data for an exemplar biosurveillance system?
4. Can an automated biosurveillance system ever really replace the astute clinician at detecting and responding to an infectious disease outbreak of public health significance?
5. What statistical approaches are available for aberration detection and what are the pros and cons of each? How does a biosurveillance system determine which aberration detection method(s) to use?
6. What are the major data quality issues related to conducting public health surveillance with ED data? How can these be identified and addressed?

7. Discuss the challenges and benefits of using secondary ED visit data for public health surveillance.
8. What are some of the security and privacy issues surrounding the use of ED visit data for public health surveillance?

REFERENCES

Aylin P, Bottle A, Majeed A. Use of Administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. BMJ 2007; 334:1044. Originally published online 23 Apr 2007.

Baer A, Jackson M, Duchin JS. What is the value of a positive syndromic surveillance signal? Advances in Disease Surveillance 2007; 2:192.

Barnett C, Deyneka L, Waller AE. Post-Katrina situational awareness in North Carolina. Advances in Disease Surveillance 2007; 2:142.

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. Pediatric Emergency Care 2004; 20(6):355–360.

Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks. MMWR. Recommendations and Reports 2004; 53(RR05):1–11.

Centers for Disease Control and Prevention. Morbidity and mortality associated with Hurricane Floyd – North Carolina, September-October 1999.MMWR. Morbidity and Mortality Weekly Report 2000; 49(17):369–372.

Centers for Disease Control and Prevention (October 23, 2003). Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp. Accessed June 24, 2005.

Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics 2001; 34(5):301–310.

Chapman WW and Dowling JN. Consultative meeting on chief complaint classifiers and standardized syndromic definitions. Advances in Disease Surveillance 2007; 4:47.

Dara J, Dowling JN, Travers D, Cooper GF, Chapman WW. Chief complaint preprocessing evaluated on statistical and non-statistical classifiers. Advances in Disease Surveillance 2007; 2:4.

Davis MV, MacDonald PDM, Cline JS, Baker EL. Evaluation of public health response to hurricanes finds North Carolina better prepared for public health emergencies. Public Health Reports 2007; 122:17–26.

Davis SF, Strebel PM, Atkinson WL, Markowitz LE, Sutter RW, Scanlon KS, Friedman S, Hadler SC. Reporting efficiency during a measles outbreak in New York City, 1991. American Journal of Public Health 1993; 83(7):1011–1015.

Espino JU, Dowling JN, Levander J, Sutovsky P, Wagner MM, Cooper GF. SyCo: A probabilistic machine learning method for classifying chief complaints into symptom and syndrome categories. Advances in Disease Surveillance 2007; 2:5.

Fleischauer AT, Silk BJ, Schumacher M, Komatsu K, Santana S, Vaz V, et al. The validity of chief complaint and discharge diagnosis in emergency-based syndromic surveillance. Academic Emergency Medicine 2004; 11(12):1262–1267.

Forbach C, Scholer MJ, Falls D, Ising A, Waller A. Improving system ability to identify symptom complexes in free-text data. Advances in Disease Surveillance 2007; 2:7.

Hakenewerth A, Waller A, Ising A, Tinitinalli J. NC DETECT and NHAMCS: Comparison of emergency department data. Academic Emergency Medicine 2009; 16(3):261–269.

Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Greenko J, et al. New York City syndromic surveillance systems. MMWR. Morbidity and Mortality Weekly Report Supplement 2004; 24(53):23–27.

Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis [serial on the Internet]. 2004 May http://www.cdc.gov/ncidod/EID/vol10no5/03–0646.htm. Accessed July 20, 2007.

Hirshon JM. For the SAEM Public Health Task Force Preventive Care project. The rational for developing public health surveillance systems based on emergency department data. Academic Emergency Medicine 2000; 7:1428–1432.

Hripscak G, Bamberger A, Friedman C. Fever detection in clinic visit notes using a general purpose processor. Advances in Disease Surveillance 2007; 2:14.

Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System. Journal of Urban Health. 2003; 80 (2 Suppl 1): i89–i96.

Ising A, Li M, Deyneka L, Barnett C, Scholer M, Waller A. Situational awareness using web-based annotation and custom reporting. Advances in Disease Surveillance 2007; 4:167.

Ising A, Travers D, Crouch J, Waller A. Improving negation processing in triage notes. Advances in Disease Surveillance 2007; 4:50.

Ising AI, Travers DA, MacFarquhar J, Kipp A, Waller A. Triage note in emergency department-based syndromic surveillance. Advances in Disease Surveillance 2006; 1:34.

Kaufman AF, Meltzer MI, Schmid GF. Economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? Emerging Infectious Diseases 1997; 3:83–94.

Komatsu K, Trujillo L, Lu HM, Zeng D, Chen H. Ontology-based automatic chief complaints classification for syndromic surveillance. Advances in Disease Surveillance 2007; 2:17.

Lee LE, Fonseca V, Brett KM, Sanchez J, Mullen RC, Quenemoen LE, Groseclose SL, Hopkins RS. Active morbidity surveillance after Hurricane Andrew--Florida, 1992. JAMA 1993; 270(5):591–594. Erratum in: JAMA 1993; 270(19):2302.

Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. Journal of the American Medical Informatics Association 2002; 9(2):105–115.

Lombardo J. A systems overview of the Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE II). Journal of Urban Health 2003; 80(2 Suppl 1):i32–i42.

Mikosz CA, Silva J, Black S, Gibbs G, Cardenas I. Comparison of two major emergency department-based free-text chief-complaint coding systems. MMWR. Morbidity and Mortality Weekly Report Supplement 2004; 53S:101–105.

National Center for Injury Prevention and Control (US). Data elements for emergency department systems, release 1.0. Atlanta, GA: Centers for Disease Control; 1997.

National Library of Medicine (2006). Specialist Lexicon and Lexical Tools Documentation, 2006AD. Retrieved May 11, 2007 from http://www.nlm.nih.gov/research/umls/meta4.html.

National Library of Medicine (2007). Unified Medical Language System Documentation, 2007AA. Bethesda, MD: National Library of Medicine. Retrieved May 11, 2007 from http://www.nlm.nih.gov/research/umls/meta2.html.

Nawar EW, Niska RW, Xu J. (2007). National Hospital Ambulatory Medical Care Survey: 2005 Emergency Department Summary. Advance data from vital and health statistics: no. 386. Hyattsville, MD: National Center for Health Statistics.

North Carolina General Statute 130A. http://www.ncleg.net/EnactedLegislation/ Statutes/ HTML/ByChapter/Chapter_130A.html Accessed January 17, 2008.

Olson DR, Heffernan RT, Paladini M, Konty K, Weiss D, Mostashari F. Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. PLoS Medicine. 2007; 4(8):e247.

Reis BY, Mandl KD. Syndromic surveillance: The effects of syndrome grouping on model accuracy and outbreak detection. Annals of Emergency Medicine 2004; 44:235–241.

Rodewald LE, Wrenn KD, Slovis CM. A method for developing and maintaining a powerful but inexpensive computer data base of clinical information about emergency department patients. Annals of Emergency Medicine 1992; 21:41–46.

Rydman RJ, Rumoro DP, Silva JC, Hogan TM, Kampe LM. The rate and risk of heat-related illness in hospital emergency departments during the 1995 Chicago heat disaster. Journal of Medical Systems 1999; 23(1):41–56.

Shapiro A. Taming variability in free text: Application to health surveillance. MMWR. Morbidity and Mortality Weekly Report 2004; 24(53S):95–100.

Sosin DM, DeThomasis J. Evaluation challenges for syndromic surveillance – Making incremental progress. MMWR. Morbidity and Mortality Weekly Report 2004; 53S:125–129.

Talan DA, Moran GJ, Mower WR, Newdow M, Ong S, Slutsker L, Jarvis WR, Conn LA, Pinner RW. EMERGEncy ID NET: an emergency department-based emerging infections sentinel network. The EMERGEncy ID NET Study Group. Annals of Emergency Medicine 1998; 32:703–711.

Teich JM, Wagner MM, Mackenzie CF, Schafer KO. The informatics response in disaster, terrorism and war. Journal of the American Medical Informatics Association 2002; 9(2):97–104.

Travers DA. (2006). Emergency Medical Text Processor. Accessed on June 10, 2008 from http://nursing.unc.edu/research/current/emtp/.

Travers DA, Barnett C, Ising A, Waller A. (2006). Timeliness of emergency department diagnoses for syndromic surveillance. Proceedings of the AMIA Symposium 2006, 769–773.

Travers DA, Haas SW. Using nurses natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. Journal of Biomedical Informatics 2003; 36:260–270.

Travers DA, Waller A, Haas S, Lober WB, Beard C. (2003). Emergency department data for bioterrorism surveillance: Electronic availability, timeliness, sources and standards. Proceedings of the AMIA Symposium 2003, 664–668.

Tsui FC, Wagner MM, Dato V, Chang C. (2001) Value of ICD-9-coded chief complaints for detection of epidemics. Proceedings of the Fall Symposium of the American Medical Informatics Association 2001, 711–715.

U.S. Department of Health and Human Services (USDHHS). International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). 6th ed. Washington: Author; 2006.

Varney SM, Hirshon JM. Update on public health surveillance in emergency departments. Emergency Medicine Clinics of North America 2006; 24:1035–1052.

Waller AE, Ising AI, Deneka L. North Carolina Biosurveillance System. In: Voeller JG, editor. Wiley Handbook of Science and Technology for Homeland Security. New York: Wiley; April 2010, vol.3.5.

Waller AE, Ising AI, Deyneka L. North Carolina Emergency Department visit data available for public health surveillance. North Carolina Medical Journal 2007; 68(4):289–291.

Wagner MM, Espino J, Tsui FC, et al. Syndrome and outbreak detection using chief-complaint data – experience of the Real-Time Outbreak and Disease Surveillance project. MMWR. Morbidity and Mortality Weekly Report 2004; 24(53S):28–31.

Weiss BP, Mascola L, Fannin SL. Public health at the 1984 Summer Olympics: the Los Angeles County experience. American Journal of Public Health 1988; s78(6):686–688.

SUGGESTED READING

Hirshon JM. For the SAEM Public Health Task Force Preventive Care project. The rational for developing public health surveillance systems based on emergency department data. Academic Emergency Medicine 2000; 7:1428–1432.

Varney SM, Hirshon JM. Update on public health surveillance in emergency departments. Emergency Medicine Clinics of North America 2006; 24:1035–1052.

Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks. MMWR. Recommendations and Reports 2004; 53(RR05):1–11.

Fleischauer AT, Silk BJ, Schumacher M, Komatsu K, Santana S, Vaz V, et al. The validity of chief complaint and discharge diagnosis in emergency-based syndromic surveillance. Academic Emergency Medicine 2004; 11(12):1262–1267.

Centers for Disease Control and Prevention (October 23, 2003). Syndrome definitions for diseases associated with critical bioterrorism-associated agents. Available at http://www.bt.cdc.gov/surveillance/syndromedef/index.asp. Accessed June 24, 2005.

Nawar EW, Niska RW, Xu J. (2007). National Hospital Ambulatory Medical Care Survey: 2005 Emergency Department Summary. Advance data from vital and health statistics: no. 386. Hyattsville, MD: National Center for Health Statistics.

ONLINE RESOURCES

Website for the International Society for Disease Surveillance (http://www.syndromic.org) This website includes the online journal Advances in Disease Surveillance, as well as a variety of wikis addressing research and public health topics.

NC DETECT website (http://www.ncdetect.org) The NC DETECT website contains links to numerous abstracts and presentations related to ED data use for biosurveillance, as well as details on the technical components of NC DETECT.