

## Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Dimitrios G. Koutsonanos

11/24/2014

---

[Student's name typed]

---

Date

PUBLIC HEALTH SURVEILLANCE SYSTEMS FOR DISEASE MONITORING,  
SITUATIONAL AWARENESS, AND DECISION MAKING SUPPORT

BY

Dimitrios G. Koutsonanos M.D.

Degree to be awarded: M.P.H.

Executive MPH

---

Paula Braun, MS  
Thesis Committee Chair

Date

---

Travis Sanchez DVM, MPH  
Thesis Field Advisor

Date

---

Melissa Alperin, MPH, MCHES  
Chair, Career MPH Program

Date

PUBLIC HEALTH SURVEILLANCE SYSTEMS FOR DISEASE MONITORING,  
SITUATIONAL AWARENESS, AND DECISION MAKING SUPPORT

BY

Dimitrios G. Koutsonanos

M.P.H., Emory University, Atlanta, GA, USA, 2014

M.D., Medical School, University of Nis, Serbia, 2004

An abstract of

a Thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements of the degree of

Master of Public Health in the Executive MPH program

2014

## **Abstract**

### **PUBLIC HEALTH SURVEILLANCE SYSTEMS FOR DISEASE MONITORING, SITUATIONAL AWARENESS, AND DECISION MAKING SUPPORT**

**BY**

**Dimitrios G. Koutsonanos**

Disease surveillance and population health monitoring represents one of the cornerstones of public health surveillance, early disease identification and prevention, and situational awareness. Constant monitoring for potential emerging public health threats is one of the most important missions of public health. Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health related event for use in public health actions to reduce morbidity and mortality and to improve health. Early detection of a new outbreak and rapid response to a public health threat could result in saving of millions of lives but also great financial savings from preventing hospitalizations, deaths and all the socio-economic effects associated with it.

In order to provide early disease detection and situational awareness, different public health surveillance systems were developed that monitor population health and report different public health events and threats. These systems perform different functions and can be classified in three different categories: i) syndromic surveillance systems, ii) laboratory surveillance systems, iii) web-based public health surveillance systems. All of these systems have significant contributions in early disease detection and monitoring, situational awareness, and decision making support.

In this study we provided an overview of major public health surveillance systems for syndromic surveillance including BioSense, ESSENCE, and ILINet, laboratory surveillance including NNDSS, FoodNet, and eHARS, and web-based surveillance including HealthMap, ProMED-mail, and BioCaster. We provided a summary of the different functionalities each system offers and supports as well as limitations associated with each system. Finally, we identified challenges in public health surveillance systems and we proposed potential tools and surveillance models that could be used to improve data collection, analysis, and reporting and enhance the functionality of public health surveillance systems.

PUBLIC HEALTH SURVEILLANCE SYSTEMS FOR DISEASE MONITORING,  
SITUATIONAL AWARENESS, AND DECISION MAKING SUPPORT

BY

Dimitrios G. Koutsonanos

M.P.H., Emory University, Atlanta, GA, USA, 2014

M.D., Medical School, University of Nis, Serbia, 2004

A Thesis submitted to the Faculty of the  
Rollins School of Public Health of Emory University  
in partial fulfillment of the requirements of the degree of  
Master of Public Health in the Executive MPH program  
2014

## **Acknowledgements**

I would like to express my deepest appreciation to my Thesis Committee Chair, Paula Braun M.S. and my Thesis Field Advisor, Travis Sanchez DVM, MPH, for their persistent help and guidance in preparing this thesis. Special thanks to the director of the Applied Public Health Informatics Program, Mark Conde for his mentorship and support during this journey.

I would like to gratefully and sincerely thank Dr. Richard W. Compans for all the opportunities and support he provided to me and Marian Compans for always making me feel welcome in their house.

Special thanks to Erin-Joi Collins McNeal for recommending the MPH program to me.

I would also like to sincerely thank Vanessa Whitehurst, Catherine Ordun, and KC Decker for giving me the opportunity to complete my practicum at Booz Allen Hamilton and be part of a great group of people.

In addition, a big thank you to all my close friends in the United States and Greece for their amazing friendship, help, and support throughout the years.

Finally, my biggest thank you and love to my mother, father, my two sisters, and their families for their unconditional love, help, and unlimited support for all the difficult decisions I have made through the years.

D.K.

Rollins School of Public Health, Emory University

PUBLIC HEALTH  
SURVEILLANCE SYSTEMS  
FOR DISEASE  
MONITORING,  
SITUATIONAL AWARENESS,  
AND DECISION MAKING  
SUPPORT

Dimitrios G. Koutsonanos. M.D.

2014

---

*Thesis Committee*

*Paula Braun M.S.  
Thesis Committee Chair*

*Travis Sanchez DVM, MPH  
Field Advisor*

*Applied Public Health Informatics Program Director  
Mark Conde*

---

## Table of Contents

<b>CHAPTER 1: OVERVIEW OF PUBLIC HEALTH SURVEILLANCE</b> .....	<b>3</b>
1.1 GOALS OF PUBLIC HEALTH SURVEILLANCE .....	4
1.2 TYPES OF PUBLIC HEALTH SURVEILLANCE.....	5
1.3 SOURCES OF DATA FOR PUBLIC HEALTH SURVEILLANCE.....	8
1.4 SCOPE OF THESIS .....	9
<b>CHAPTER 2: SYNDROMIC SURVEILLANCE SYSTEMS</b> .....	<b>11</b>
2.1 BIOSENSE .....	14
2.2 ESSENCE.....	18
2.3 ILINet.....	23
2.4 OTHER SYNDROMIC SURVEILLANCE SYSTEMS .....	25
<b>CHAPTER 3: LABORATORY SURVEILLANCE SYSTEMS</b> .....	<b>26</b>
3.1 NNDSS.....	26
3.2 FOODNET .....	31
3.3 eHARS .....	33
3.4 OTHER LABORATORY SURVEILLANCE SYSTEMS .....	34
<b>CHAPTER 4: WEB-BASED PUBLIC HEALTH SURVEILLANCE SYSTEMS</b> .....	<b>35</b>
4.1 HEALTHMAP .....	37
4.2 PROMED-MAIL.....	40
4.3 BIOCASTER.....	42
4.4 OTHER WEB-BASED SURVEILLANCE SYSTEMS.....	44
<b>CHAPTER 5: THE FUTURE OF PUBLIC HEALTH SURVEILLANCE SYSTEMS</b> .....	<b>45</b>
5.1 PREDICTIVE ANALYTICS IN PUBLIC HEALTH.....	45
5.2 ELECTRONIC HEALTH RECORDS.....	48
5.3 NoSQL DATABASES .....	49
5.4 DESIGNING THE IDEAL PUBLIC HEALTH SURVEILLANCE SYSTEM.....	50
<b>REFERENCES</b> .....	<b>57</b>



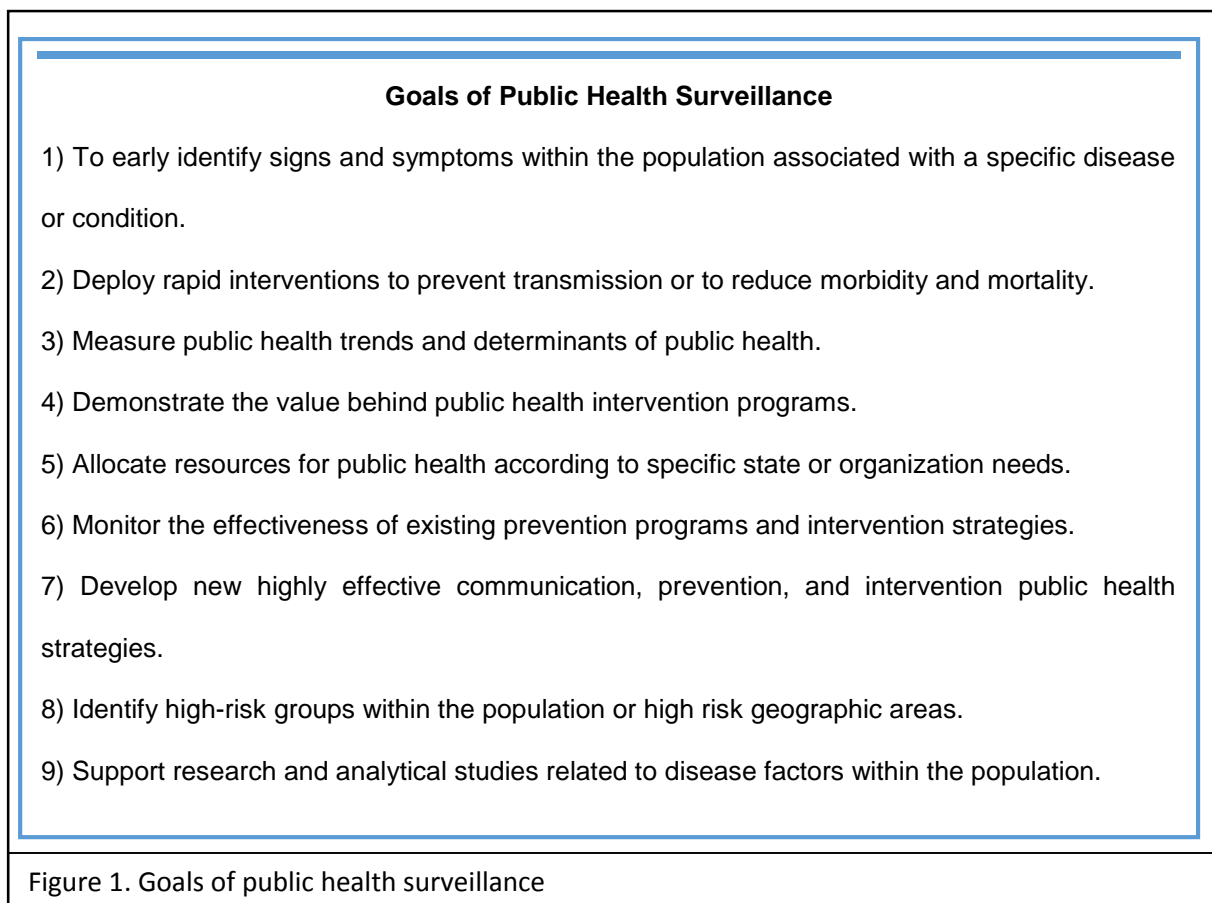
## **CHAPTER 1: OVERVIEW OF PUBLIC HEALTH SURVEILLANCE**

Disease surveillance represents one of the cornerstones of public health surveillance, early disease identification and prevention, and situational awareness (Adams et al., 2014; L. M. Lee, Thacker, Centers for Disease, & Prevention, 2011; Thacker, Qualters, Lee, Centers for Disease, & Prevention, 2012). Constant monitoring for potential emerging public health threats is one of the most important missions of public health ("Addressing emerging infectious disease threats: a prevention strategy for the United States. Executive summary," 1994). Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health related event for use in public health action to reduce morbidity and mortality and to improve health (Buehler, Centers for Disease, & Prevention, 2012; Smith et al., 2013; Thacker et al., 2012). Early detection of a new outbreak could result in saving of millions of lives but also great financial savings from preventing hospitalization, deaths and all the socio-economic effects associated with it. Public health surveillance is based on the ongoing, systematic collection and analysis of health-related data. Based on this analysis, data will be interpreted and further support decision making, strategic planning, and evaluation of existing public health practices and programs for disease control and prevention (Nsubuga et al., 2006). Public health surveillance goals include the assessment of population health, the identification of public health priorities, the evaluation of existing public health programs, and the development of new effective interventions and strategies to protect and improve public health (Nsubuga et al., 2006; Thacker, Berkelman, & Stroup, 1989). Public health surveillance covers all aspect of public health threats and diseases, such as communicable and non-communicable

diseases, environmental health, toxic exposures, natural and man-made disasters, and health behaviors (2007; Alwan et al., 2010; Hay, George, Moyes, & Brownstein, 2013; Heymann & Rodier, 1998). Public health surveillance provides an estimate about the magnitude of the problem, the characteristics of the affected population, the geographical distribution of the condition, and characteristics and the natural history of the condition that impact the population.

### **1.1 GOALS OF PUBLIC HEALTH SURVEILLANCE**

Public health surveillance is being used from National agencies in order to monitor disease trends over time to support policy changes and implementation of specific



regulations but also to protect the public from potential biological threats. Public health surveillance is also being used by State and local health departments and agencies for

managing population health exposures, assuring accurate diagnosis and treatments of affected populations, detect outbreaks, and guide public health prevention and control programs. According to the Centers for Disease Control and Prevention (CDC) and the blueprint for a national public health surveillance system for the 21<sup>st</sup> century, the goals of public health surveillance are summarized into Figure 1 (Smith et al., 2013).

## **1.2 TYPES OF PUBLIC HEALTH SURVEILLANCE**

In order to protect and improve public health, achieve early disease detection and situational awareness, several surveillance systems have been developed that monitor and register the appearance of signs and symptoms among the general population, record laboratory results and clinical diagnosis, but also monitor social behaviors and social events in order to indicators of public health interest and potential health risks (Berkelman & Buehler, 1990; Buehler et al., 2012; Choi, 2012; Howard, 2000; Jajosky & Groseclose, 2004; Morse, 2012; Wojcik, Brownstein, Chunara, & Johansson, 2014). According to these specifications, different types of public health surveillance systems have been developed. Among these, three distinct types of surveillance systems with significant public health functions can be identified based on reporting requirements as shown in figure 2; syndromic surveillance, laboratory surveillance, and web-based surveillance systems.

Syndromic surveillance systems have been designed as early event detection and situational awareness systems (Buehler et al., 2009; Samoff, Fangman, Hakenewerth, Ising, & Waller, 2014; Uscher-Pines et al., 2009). In order to achieve these goals, their

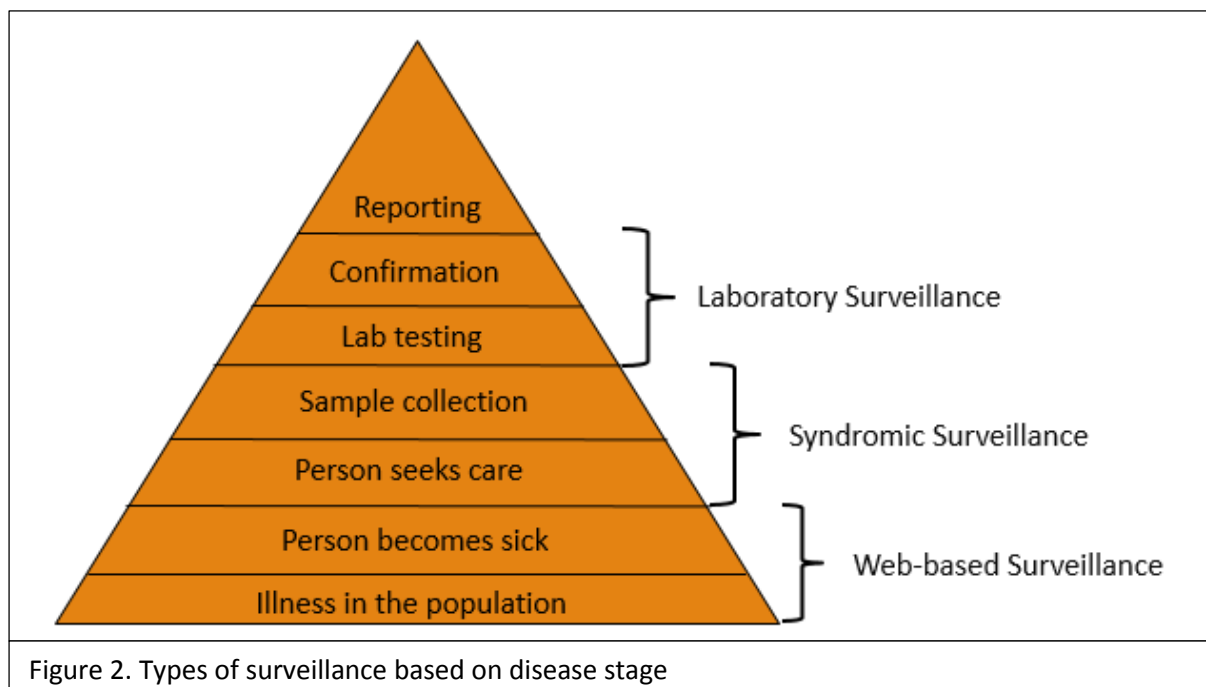


Figure 2. Types of surveillance based on disease stage

general framework was based on the collection of pre-diagnostic data, such as chief symptoms and complaints. By grouping these symptoms and complains in different sub-syndromes and syndromes, the potential etiological factor behind the reported condition could be identified and thus provide early recognition of emerging trends for the community being monitored (Betancourt, Hakre, Polyak, & Pavlin, 2007; Buckeridge et al., 2004; Chen et al., 2005; Pattie, Atherton, & Cox, 2009; Reis & Mandl, 2004).

Currently, there are several systems used by different states and jurisdictions in the United States for monitoring population health. Some of the most widely used systems with high adoption rates from local and state health departments include BioSense, the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), and the Influenza-Like Illness Network (ILINet) (Buehler et al., 2009; Espino et al., 2004; Fricker, Hegler, & Dunfee, 2008; Howard, 2000; Lazarus, Kleinman, Dashevsky, DeMaria, & Platt, 2001; Lemay, Mawudeku, Shi, Ruben, & Achonu, 2008; J. S. Lombardo, Burkom, & Pavlin, 2004; Morse, 2012; Xing, Burkom, &

Tokars, 2011). Syndromic surveillance systems provide a rapid, near real time method for disease monitoring and reporting.

Laboratory surveillance systems are based on data generated and reported from clinical and/or public health laboratories (McElwain, 2010). Laboratory based surveillance monitors for a range of food and waterborne pathogens, sexually transmitted and blood-borne diseases, respiratory pathogens, zoonotic diseases, toxins and environmental pollutants, as well as non-communicable diseases and conditions such as diabetes, hypercholesterolemia, cancer markers, etc (Henao, Crim, & Hoekstra, 2012; MacIntosh, Tastad, & Eick-Cost, 2013; Respass, Rayfield, & Dondero, 2001; Richardson et al., 2010; Sintchenko & Gallego, 2009). Furthermore, since laboratory results are highly standardized based on well-established protocols and standard operating procedures, laboratory surveillance can represent a sensitive method for disease and condition monitoring (Bronnert et al., 2014; Hunscher, Boyd, Green, & Clauw, 2006; L. H. Lee, Gross, Hartung, Liou, & Rahm, 2014; Mok, Ho, Tsui, Ng, & Fung, 2013; Paraiso, Perez Del Rey, Bucur, Claerhout, & Alonso-Calvo, 2014; Ranallo et al., 2013). In contrast to syndromic surveillance systems, laboratory surveillance systems are based on laboratory confirmed results that can be traced back to the individual patient level and can allow direct patient intervention and contact tracing since all laboratory samples and specimens are linked to the individual using unique identifiers. Some of the most important laboratory surveillance systems with critical public health roles include CDC's National Notifiable Diseases Surveillance System (NNDSS) and National Electronic Disease Surveillance System (NEDSS), Foodborne Diseases Active Surveillance Network (FoodNet), and CDC's National HIV Surveillance System. (Cohen, Gray,

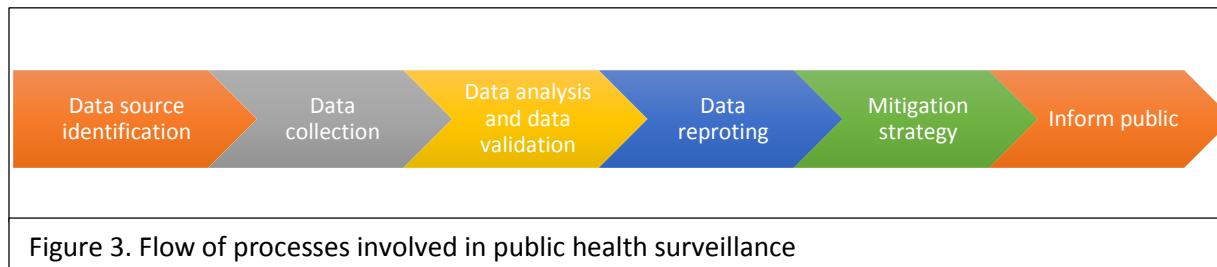
Ocfemia, Johnson, & Hall, 2014; "Communicable diseases surveillance. Presentation of NNDSS data," 2001; Hurd et al., 2012; Jajosky & Groseclose, 2004; National Electronic Disease Surveillance System Working, 2001; Paz-Bailey, Raymond, Lansky, & Mermin, 2014; Scallan & Mahon, 2012).

Web-based surveillance systems have been designed to use near real-time information from the internet, social media, news media, and reports and by applying filtering methods, natural language processing algorithms, and machine learning techniques, identify signs and symptoms reported by individuals and may be associated with specific diseases and conditions (Brownstein, Freifeld, & Madoff, 2009; Milinovich, Williams, Clements, & Hu, 2014). These systems can search through millions of online sources, internet posts, reports, and blogs in multiple languages and geographic regions and identify individual cases or cluster of cases that indicate the presence of a disease or condition. HealthMap and the Program for Monitoring Emerging Diseases (ProMED-mail) are the most widely used and well recognized web-based surveillance systems due to identification or recent outbreaks such as the Ebola and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Other web-based surveillance systems include Biocaster, Global Public Health Intelligence Network (GPHIN), and others (Freifeld, Mandl, Reis, & Brownstein, 2008; Lyon, Nunn, Grosse, & Burgman, 2012; Madoff, 2004; Madoff & Woodall, 2005; Morse, 2012).

### **1.3 SOURCES OF DATA FOR PUBLIC HEALTH SURVEILLANCE**

Several of the described public health surveillance systems can use multiple sources for health-related data. According to the World Health Organization (WHO) these sources can include: mortality reports, morbidity reports, epidemic reports, laboratory reports,

reports of individual case investigations, reports of epidemic investigations, special surveys (e.g., hospital admissions, disease registers, and serologic surveys), information on animal reservoirs and vectors, demographic data, and environmental data (Buckeridge et al., 2004; Buehler et al., 2004; Morse, 2012; Sosin, 2003a). In addition, alternative sources of data can include emergency-department visit data, intensive care unit data, outpatient visits, hospital admission and discharge systems, illness-related 911 calls, pre-diagnostic laboratory data (ICD-9 codes), over-the-counter medication purchases, social data (school and work absenteeism), nurse hotline calls and others (T. Andersson et al., 2014; Betancourt et al., 2007). Finally, unconventional sources for health related data include social media such as Facebook and Twitter, RSS feeds, blogs, as well as media and news reports (Hay et al., 2013; Milinovich et al.,



2014). All these sources can directly or in-directly provide health-related data. In the United States, all these conventional and alternative data sources are being used to support public health surveillance systems and public health surveillance efforts. A high level overview of the processes from data source identification to data analysis and response to a condition is illustrated in Figure 3.

#### **1.4 SCOPE OF THESIS**

Here we provided an overview of the functionality of the major surveillance systems used in syndromic, laboratory, and web-based surveillance in the United States. To

achieve this, we used a multidimensional approach and collected relevant information based on literature reviews, published data, reports and evaluations of these systems, as well as information gained from functional communities, user guides, and information provided by the official pages of each surveillance system. In addition, we collected relevant information from state and local health departments but also from federal sources (CDC, HHS), the Association of State and Territorial Health Officials (ASTHO), the Council of State and Territorial Epidemiologists (CSTE), the National Association of County and City Health Officials (NACCHO) and the International Society for Disease Surveillance (ISDS). Collecting data from multiple sources allowed us to locate the most updated and accurate information about public health surveillance systems currently used in different local and state health departments and organizations throughout the US. Finally, we identified challenges and limitations associated with different types of surveillance systems and proposed tools, approaches, and methodologies that can be used to improve the effectiveness of existing surveillance systems or design new and highly effective public health surveillance systems. The materials and methods used for this study including literature sources, reports, and keywords are shown in figure 4.

<b>MATERIALS AND METHODS</b>	
Literature search	<ul style="list-style-type: none"> <li>▪ PubMed</li> <li>▪ Google Scholar</li> <li>▪ Web of Science</li> </ul>
Keywords	Health surveillance, syndromic surveillance, laboratory surveillance, BioSense, ESSENCE, RODS, NNDSS, NEDSS, FoodNet, ILInet, HealthMap, ProMED-mail, Biocaster, detection algorithm, disease detection, NLP, NoSQL, disease predictive models, predictive analytics, electronic health records
Other sources	Official reports, federal agencies (HHS, CDC, FDA, etc.), official sites, ASTHO, CSTE, NACCHO, local and state health departments

Figure 4. Materials and Methods sources and collection methodology



## **CHAPTER 2: SYNDROMIC SURVEILLANCE SYSTEMS**

It is widely accepted that syndromic surveillance is one of the foundations of public health prevention (Dorea, Lindberg, McEwen, Revie, & Sanchez, 2014; Kaufman & Shohat, 2014; Samoff et al., 2014; Wojcik et al., 2014). Early disease identification and situational awareness are crucial elements that can support decision making and rapid, well organized, and effective response to public health threats. These approaches have a direct public health impact by protecting and saving millions of lives and additionally an economic impact by reducing financial loss associated with disease hospitalization, outpatient visits, life-years loss and the overall disease burden (Katz, May, Baker, & Test, 2011; O'Connell, Zhang, Leguen, Llau, & Rico, 2010). Syndromic surveillance has been developed as a strategy for early event detection, situational awareness and monitoring of public health. In order to achieve these goals, it was designed with a general framework of collecting pre-diagnostic data, such as chief symptoms and complaints (i.e. coughing, fever) and by grouping them in different sub-syndromes and syndromes (respiratory, gastrointestinal, etc.), match them with potential etiological factors and pathogens and identify emerging trends that appear in the community being monitored (Buckeridge et al., 2004; Buehler et al., 2004; Centers for Disease & Prevention, 2002; Guasticchi et al., 2009; Sosin, 2003b). Because these systems are based on syndrome monitoring and reporting, before formal diagnoses is made, they were named syndromic surveillance systems. The effectiveness of different syndromic surveillance systems for accurate disease identification trends and decision making support depends on several factors such as the number of providers contributing data to the system, the number of patients and reports the system captures, the geographic

distribution /coverage of providers and patients, the frequency these sources submit data to the system, the quality of reported data (completeness of data, duplication of data), the type and sensitivity of the algorithms these systems use for disease detection, the frequency reported data are being analyzed by system users, and finally the use of automated mechanisms or tools for data reporting, data validation and data analysis (Buehler et al., 2004; Dorea et al., 2014; Karami, 2012; Kashiouris, O'Horo, Pickering, & Herasevich, 2013).

The development of syndromic surveillance systems became essential as a result of the 2001 anthrax terrorist attacks and the realization that a real-time or near real-time alarming system that will monitor population health and provide situational awareness was needed (Buehler et al., 2009; Centers for Disease & Prevention, 2002; Nordin et al., 2005; van den Wijngaard, van Pelt, Nagelkerke, Kretzschmar, & Koopmans, 2011). These systems should be able to provide notification earlier than conventional /traditional surveillance systems that are based on disease case and laboratory results reporting. In order to achieve this, these systems collect data related to early states and clinical manifestation of a disease and by further analyzing these syndromes and

frequency of reported complaints to predict the emerge of diseases or conditions occurring in the population. The

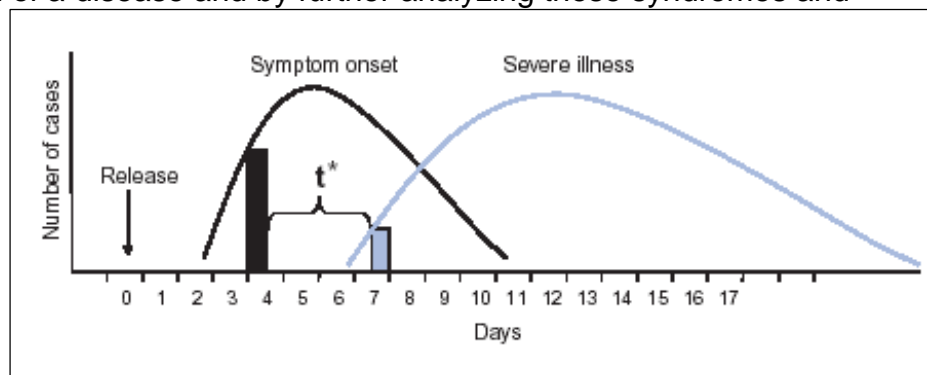


Figure 5. Time between detection by syndromic (pre-diagnostic) surveillance and detection by traditional (diagnoses-based) surveillance  
 \*Source CDC, MMWR: Sep.24, 2004/ 53(Suppl);5-11

concept behind syndromic surveillance and reporting of disease syndromes and symptoms for early disease detection is illustrated in figure 5.

Since their inception, these systems have advanced and are currently used not only for the detection of bioterrorism attacks, but also for the recognition of emerging infectious disease and other conditions that represent a public health threat. There are currently multiple systems used for syndromic surveillance including BioSense, ESSENCE, and ILINet with significant operational differences related to sensitivity, specificity, sources of syndromic data, and detection algorithms. Multiple public health sources can support the function of these systems by providing different types of syndromic surveillance data in real-time, emergency-department visits, intensive care units, outpatient visits, hospital admission and discharge systems, illness-related 911 calls, pre-diagnostic laboratory data (ICD-9, ICD-10 codes), over-the-counter medication purchases, social data (school and work absenteeism), nurse hotline calls and others (T. Andersson et al., 2014; Betancourt et al., 2007; Sugawara et al., 2007; Velardi, Stilo, Tozzi, & Gesualdo, 2014). All these data can have indicators of possible public health events occurring in different populations and different geographic locations and could infer patterns suggestive of an outbreak. All these syndromic surveillance systems need to perform 4 basic functions: 1) early disease detection in the population and rapid reporting, 2) rapid analysis our collected data and data cross validation, 3) support decision making and response support, and 4) provide support for evaluating current interventions and support health care policies and programs. In this chapter we will provide a description, functionality, differences, and public health role of the most widely used syndromic surveillance systems currently used for public health purposes.

## **2.1 BIOSENSE**

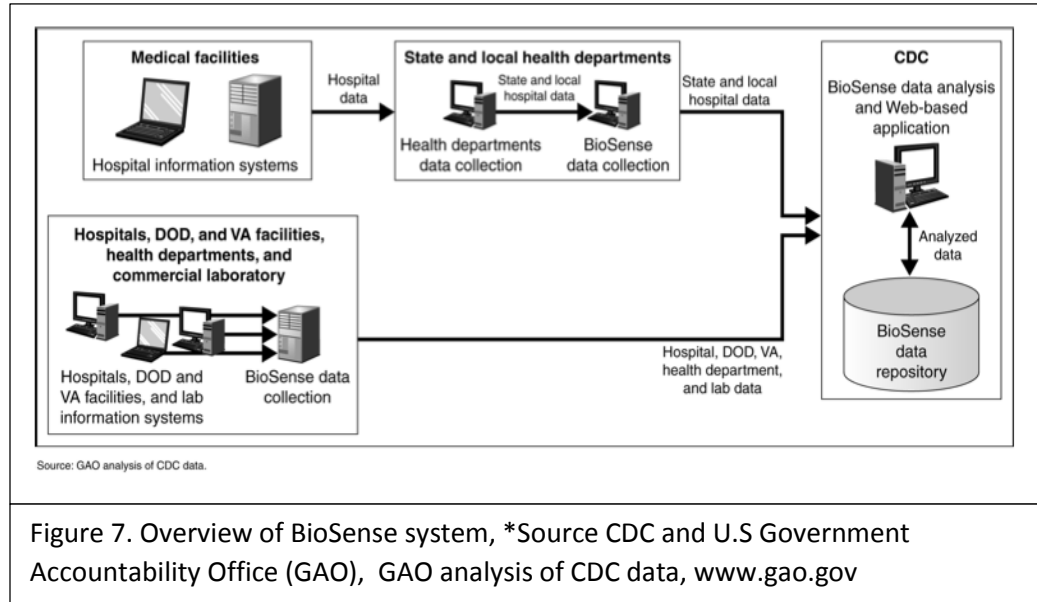
BioSense is the main CDC effort in supporting early disease detection against potential biological threats and other events of public health concern on a national level (Bradley, Rolka, Walker, & Loonsk, 2005; Loonsk, 2004; Ma et al., 2005). The original BioSense system was first developed in 2003-2004 and incorporated data from three different data sources, the Department of Defense, the Department of Veterans Affairs, and laboratory data from the Laboratory Corporation of America (LabCorp) (Bradley et al., 2005; Loonsk, 2004; Sokolow et al., 2005). In 2008 BioSense was redesigned in order to collect and analyze data from state and local health departments, health care facilities and hospitals, outpatient pharmacy data, and other laboratory sources such as Quest Diagnostics. The sources the current BioSense iteration uses to collect health related data and related reporting requirement timelines are shown in figure 6.

<b>BioSense Data Sources</b>	
Civilian Hospitals	~640 facilities (~12% ED coverage in US, Chief complains: median 24 hour latency, Diagnoses: median 6 days latency)
Veterans Affairs and Department of Defense	~1400 facilities in 50 states and DC (final diagnosis 2-5 days latency)
National Labs (LabCorp and Quest)	~47 states and DC (24 hour latency)
Hospital Labs	~49 hospital labs in 17 states/jurisdiction (24 hour latency)
Pharmacies	~50.000 in 50 states (24 hour latency)

Figure 6. BioSense data sources and reporting timelines, \*Source CDC, <http://www.cdc.gov/biosense/index.html>

An overview of the current BioSense system architecture is illustrated in figure 7 (Bradley et al., 2005; Tokars, English, McMurray, & Rhodes, 2010).

BioSense categorizes reported data into eleven different groups /syndrome categories



including botulism-like, fever, gastrointestinal, hemorrhagic illness, localized cutaneous lesion, lymphadenitis, neurological, respiratory, rash, severe illness or death, and specific infection as shown

in figure 8 (Tokars et al., 2010; Xing et al., 2011).

These 11 syndromes were defined in 2003 by a multi-agency working group to contain the prodromes of infectious diseases

BioSense Syndrome Groups	
Botulism-like	Fever
Gastrointestinal	Hemorrhagic illness
Localized cutaneous lesion	Lymphadenitis
Neurological	Respiratory
Rash	Severe illness or death
Specific infection	

Figure 8. BioSense syndromic group categorization

potentially caused by bioterrorism. Based on these and categories and information from existing systems, 78 sub-syndromes were defined to allow for analysis of more specific symptoms such as cough, asthma, difficulty breathing, etc. and to account for infectious diseases, chronic diseases, injuries, and exposures. As a result, BioSense provides the

ability to state and local health departments to perform syndromic surveillance within their own jurisdictions and share collected data with all interested stakeholders.

BioSense has evolved into a syndromic surveillance system that provides:

- A Meaningful Use-ready, HIPAA-compliant environment that provides a “Catcher’s Mitt” where health departments can receive and store automated, health-related data.
- Analytics tools that jurisdictions can use for data analysis within the BioSense infrastructure.
- A shared space for easy data exchange with other jurisdictions on an ad hoc or routine basis as selected by users.
- A method to accept electronic data in many forms and formats, including HL7, as well as convert these data to new formats for easier health department use.

Upon data reception and assignment into one of the 11 main syndrome groups and 78 sub-syndromes, disease detection is done using a modified version of the C2 algorithm for signal analysis. There are 3 main disease detection algorithms used in disease surveillance systems, C1, C2, and C3. These algorithms can establish a functional baseline (background) based on analyzed data and detect signals that differentiate/ deviate from this baseline. Briefly, the C1, C2, and C3 algorithms were intended to be CUSUM-like methods (Fricker et al., 2008; O'Brien & Christie, 1997; Zikos & Diomidous, 2012). Though, the C1 and C2 are actually Shewhart procedure variants that use a moving sample average and sample standard deviation to standardize each observation. Based on these parameters, the sensitivity of each algorithm differs. The C1 algorithm uses the seven days prior to the current observation to calculate the

sample average and sample standard deviation. The C2 is similar to the C1 but uses the seven days prior to a two-day lag. The C3 algorithm combines information from C2 statistics as described below. C1 is the less sensitivity, with C3 the most sensitive, and C2 with intermediate sensitivity. BioSense uses a modified version of the C2 algorithm named W2 for disease/anomaly detection. This algorithm uses a sliding baseline of 7 consecutive recent days' counts to calculate a mean ( $\mu$ ) and SD ( $s_t$ ). The test statistic is  $(x_t - \mu)/s_t$ , the number of SDs by which the current value  $x_t$  exceeds  $\mu$ , or 0 if  $x_t$  does not exceed  $\mu$ .

Once the data gets analyzed, they can be visualized in the main BioSense user interface that demonstrates the geographic location of the data/cases, as well as a histogram of reported cases compared to base line. More advanced analytical tools can be used for further analysis (Benoit, McDonald, English, & Tokars, 2011; Buehler et al., 2009; Xing et al., 2011). The main user interface for data visualization can be seen in

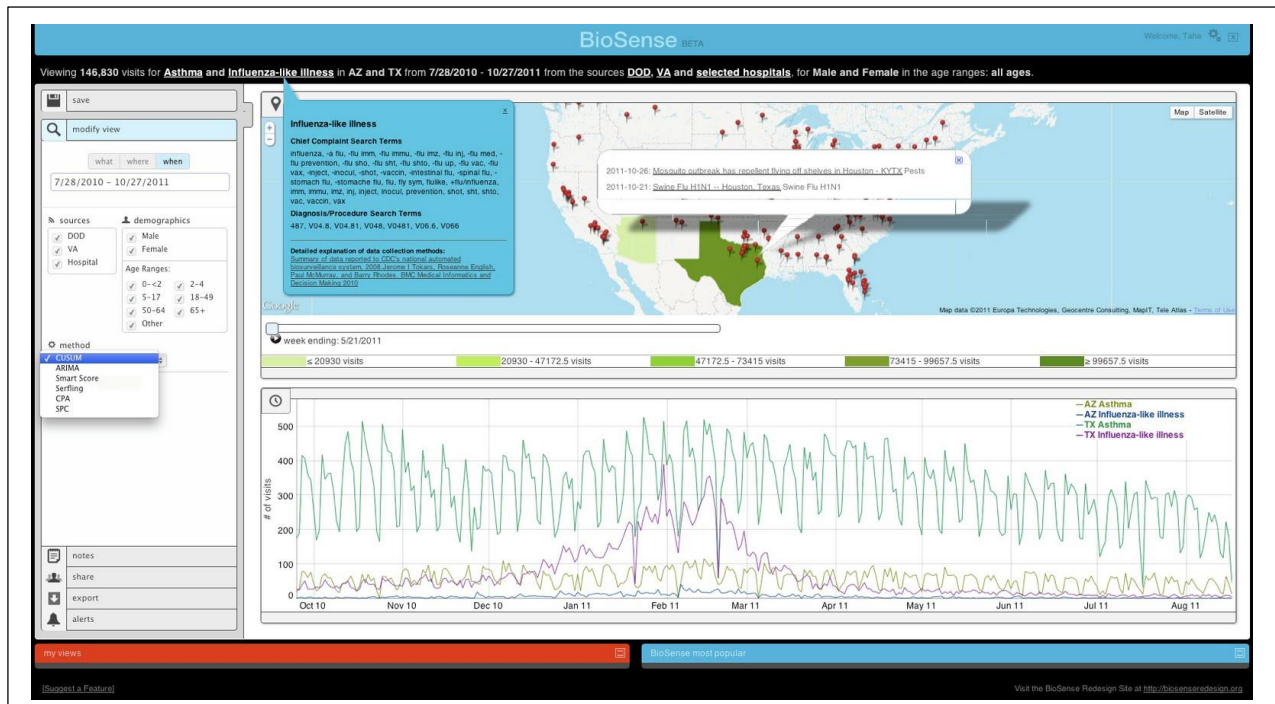


Figure 9. Data visualization in BioSense, \*Source CDC, <http://www.cdc.gov/biosense/action.html>

figure 9 and includes results of the analysis, case counts, time series graphs, geolocation, and detailed case reports.

BioSense represents the first Department of Health and Human Services system to move completely to a distributed cloud computing environment governed jointly by local, state, and federal public health representatives and one of the largest currently operational syndromic surveillance (Bradley et al., 2005; Loonsk, 2004). BioSense program aims to become the main syndromic surveillance system used by state and local health departments and jurisdictions and be used as a tool for early disease detection or conditions circulating in the population, and improve public health situational awareness and support rapid decision making. Since its inception though, BioSense has faced several challenges with most important relating to data collection and reporting at the local and state level, delays with data transport to CDC, and delays in data analysis (Buehler et al., 2009; Sokolow et al., 2005).

## **2.2 ELECTRONIC SURVEILLANCE SYSTEM FOR THE EARLY NOTIFICATION OF COMMUNITY-BASED EPIDEMICS (ESSENCE)**

The Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) represents one of the oldest, most mature syndromic surveillance systems built and currently used by multiple states and public health departments nation-wide (Holtry, Hung, & Lewis, 2010). ESSENCE development started in 1999 as a collaboration between the Johns Hopkins University Applied Physics Laboratory (JHU/APL) and Dr. Joseph Lombardo, and the Walter Reed Army Institute of Research (WRAIR) under the sponsorship of the Defense Advanced Research Projects Agency (DARPA). The system was intended to be used in the Department of Defense



Global Emerging Infections System (DoD-GEIS) (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004).

ESSENCE was a revolutionary system that combines nontraditional health status indicators and data with new analytical approaches to identify abnormal health conditions in the population. ESSENCE is constructed by multiple components developed in parallel and linked under a unified platform. The multi-component architecture allows for different components to be upgraded in a nonproprietary environment with other cost effective modules or software (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004). According to JHU/APL, ESSENCE was developed as a syndromic surveillance system that can provide the following functionalities:

- Policies to ensure the privacy of personal health care information
- Policies to govern the exchange of information among other surveillance or reporting systems
- A data archive
- Processes for detection of and issuing alerts about abnormalities in the indicator data
- Processes for notification of users of special events or environmental conditions that warrant changes in detection parameters
- Processes that allow the user to exploit the archive fully to identify false positives or obtain information about current or historical trends in the indicator data
- Visualization and user interfaces
- Processes for injecting simulated data for training and measuring the performance of ESSENCE detectors and indicators

A diagram of ESSENCE's architecture is illustrated in figure 10 (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004).

In contrast to BioSense, ESSENCE classifies all reported conditions into 7 syndrome groups including, respiratory, gastrointestinal, rash,

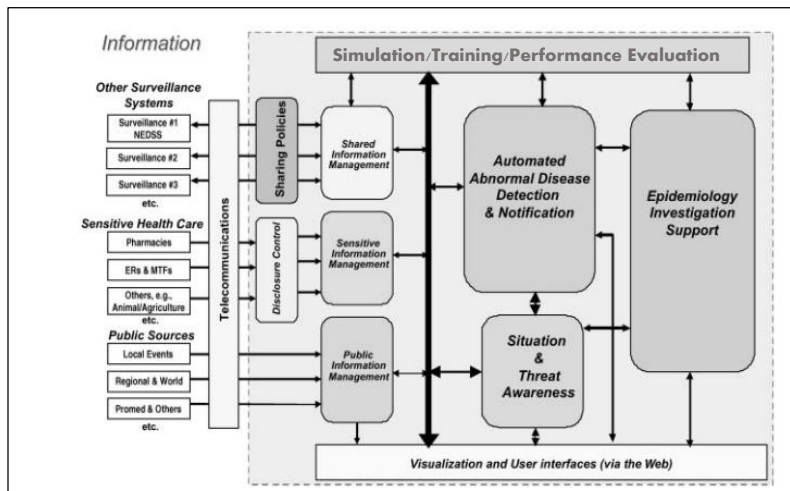


Figure 10. Overview of ESSENCE system, Lombardo et al. 2004

shock/coma, neurological, hemorrhagic, and Botulism-like as shown in figure 11 (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004).

ESSENCE collects data from sources that can be categorized in three groups: sensitive health care information, publicly available information, and products of external surveillance. Under these three categories, a variety of data can be collected including Emergency Department (ED) chief complaints (ICD-9, ICD-10 codes), over-the-counter

ESSENCE Syndrome Groups	
Botulism-like	Gastrointestinal
Neurological	Hemorrhagic
Respiratory	Rash
Shock/coma	

Figure 11. ESSENCE syndromic grouping

(OTC) pharmaceutical sales, 911 calls, nurse hotline calls, poison control center calls, visits to private practice physicians and military clinics, requests for laboratory work, laboratory results, emergency room visits, prescription medications, as well as environmental and weather data as shown in figure 12 (Betancourt et al., 2007; Henry, Magruder, & Snyder, 2004; Holtry et al., 2010).

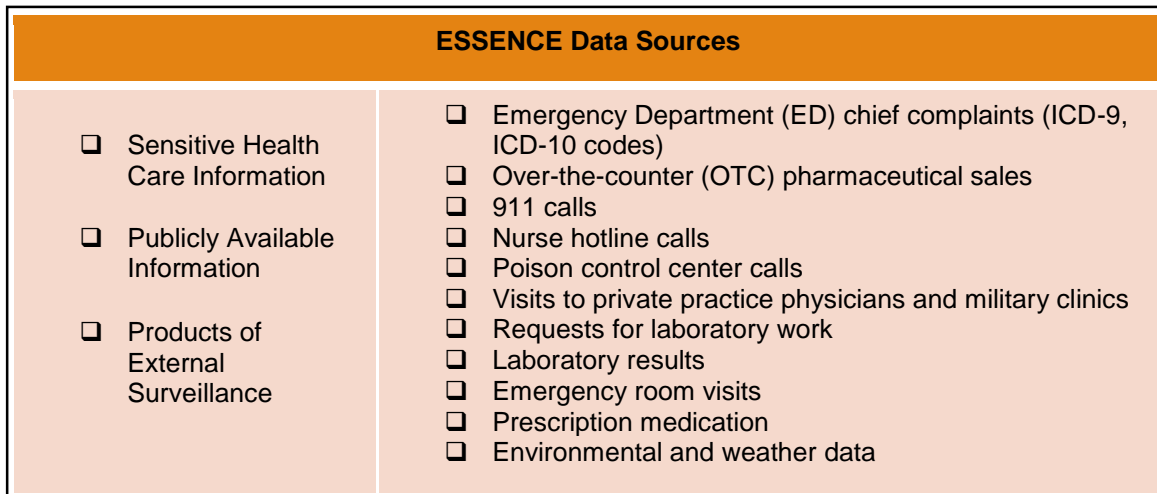


Figure 12. ESSENCE data sources

ESSENCE is capable of performing automatic analysis of reported data against a baseline that is established based on a 28-day average and identify statistically significant changes. As a result, ESSENCE will “flag” a syndrome group that falls outside the baseline measurement defined by historical data. Using this approach, ESSENCE can provide early warning for anomalies that are detected among the general population (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004).

ESSENCE allows users in local and state health departments to define the sensitivity of alerts depending of public health needs and current conditions by selecting one of the three

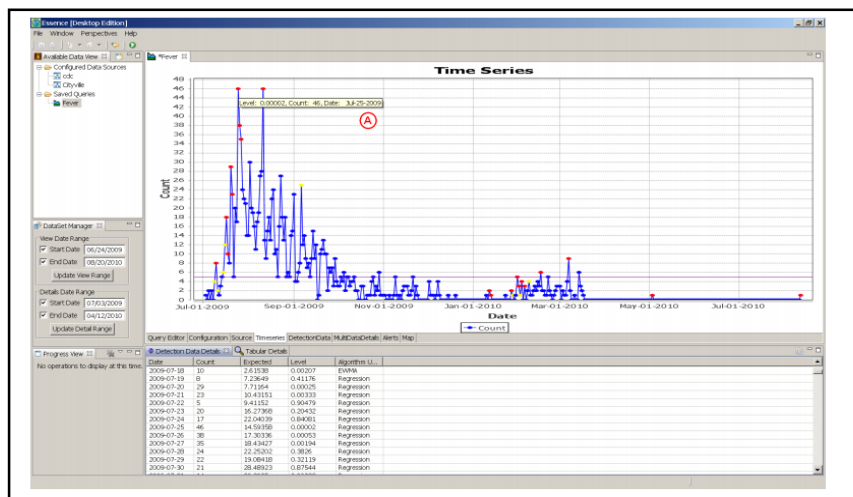


Figure 13. Overview of ESSENCE analytical interface, Lombardo et al. 2004

detection algorithms C1, C2, and C3. In addition, ESSENCE provides additional algorithms and automatically selects between a regression-based algorithm and an

adaptive Exponentially Weighted Moving Average (EWMA) chart, with the selection determined by a goodness-of-fit measure for the regression model (Carnevale et al., 2011; Dorea, McEwen, McNab, Sanchez, & Revie, 2013; Fricker et al., 2008; Zikos & Diomidous, 2012). Based on this algorithms, ESSENCE can display a time series analysis graph of the results and other advanced analytical options as seen in figure 13 (J. Lombardo et al., 2003; J. S. Lombardo et al., 2004; Savory, Cox, Emch, Alemi, & Pattie, 2010).

Finally, ESSENCE provides the option for geospatial analysis and data visualization (cluster analysis)

as illustrated in figure 14 (Holtry et al., 2010; J. S. Lombardo et al., 2004).

Overall, ESSENCE has been one of the

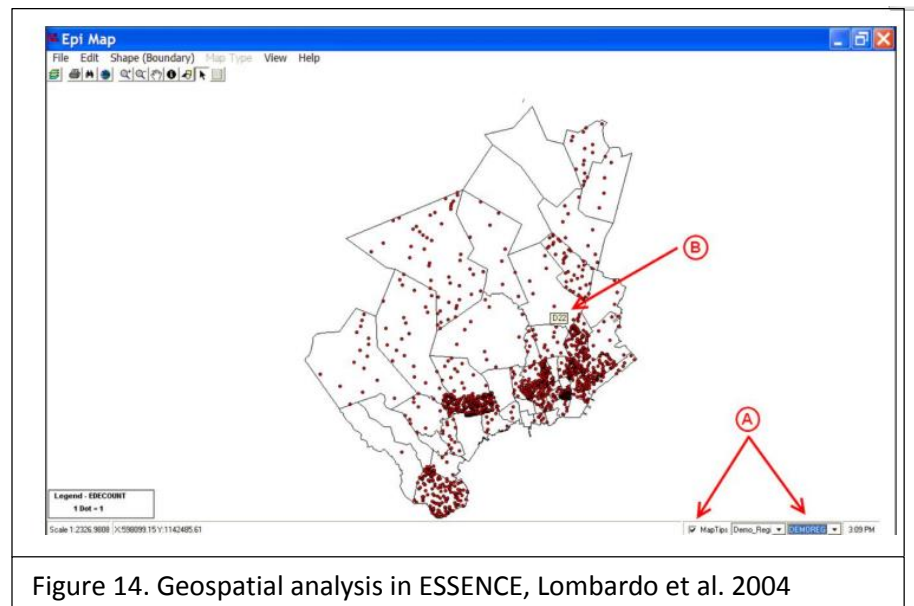


Figure 14. Geospatial analysis in ESSENCE, Lombardo et al. 2004

most successful syndromic surveillance systems with high rates of adoption from several state and local health departments and positive results regarding early disease detection, notification, and protection of public health (Centers for Disease & Prevention, 2011; Holtry et al., 2010; Schirmer, Lucero, Oda, Lopez, & Holodniy, 2010).

### **2.3 INFLUENZA-LIKE ILLNESS NETWORK (ILINet)**

The Influenza-Like illness network (ILINet) is a syndromic surveillance system that exclusively focuses in monitoring and reporting influenza activity in the United States (Scarpino, Dimitrov, & Meyers, 2012). ILINet is a national outpatient influenza illness surveillance program where healthcare providers that participate in the network report data related to patient visits with symptoms associated with influenza illness (Patwardhan & Bilkovski, 2012; Scarpino et al., 2012). For ILINet surveillance the following case definition for influenza-like illness (ILI) is used: **Fever (>100°F or >37.8°C) and cough and/or sore throat without a known cause other than influenza** (Centers for Disease & Prevention, 2010; Scarpino et al., 2012). According to this case definition, ILINet will collect all cases report by network providers and we categorize them in five different age groups: 0-4 years, 5-24 years, 25-49 years, 50-64 years, and 65+years of age. Network participants provide weekly reports about

influenza cases directly to CDC. Besides reports, ILINet providers can additionally submit patient samples for laboratory characterization. Medical providers of any specialty in

<b>ILINet Data Sources</b>	
Medical Providers	Pediatric Facilities
Emergency Departments	Student Health Care Centers
Family Practices	Urgent Care Centers
Internal Medicine Clinics	Geriatric Facilities
Infectious Disease Clinics	OB/GYN

Figure 15. ILINet data source provides

any facility are eligible to become ILINet providers including Emergency Medicine departments, family practices, infectious disease clinics, internal medicine clinics, OB/GYN, pediatric facilities, student health centers, and urgent care centers as shown

in figure 15. Currently there are more than 2,900 outpatient health care providers in all 50 states and the District of Columbia that participate in the network and report more than 30 million patient visits annually (Scarpino et al., 2012).

ILINet collects reported data and the percentage of patient visits to healthcare providers for ILI reported each week is weighted on the basis of state population. This percentage is compared each week with the national baseline of 2.0% calculated by the mean percentage of patient visits for ILI during non-influenza weeks for the previous three seasons and adding two standard deviations (Centers for Disease & Prevention, 2010). A non-influenza week is defined as periods of two or more consecutive weeks in which each week accounted for less than 2% of the season's total number of specimens that tested positive for influenza. Due to wide variability in regional level data, it is not appropriate to apply the national baseline to regional data; therefore, region specific baselines are calculated using the same methodology. Based on this analysis, ILINet reports influenza activity throughout the US in all age groups as illustrated in figure 16

(Centers for Disease & Prevention, 2010; Patwardhan & Bilkovski, 2012). As a result, ILINet allows for monitoring

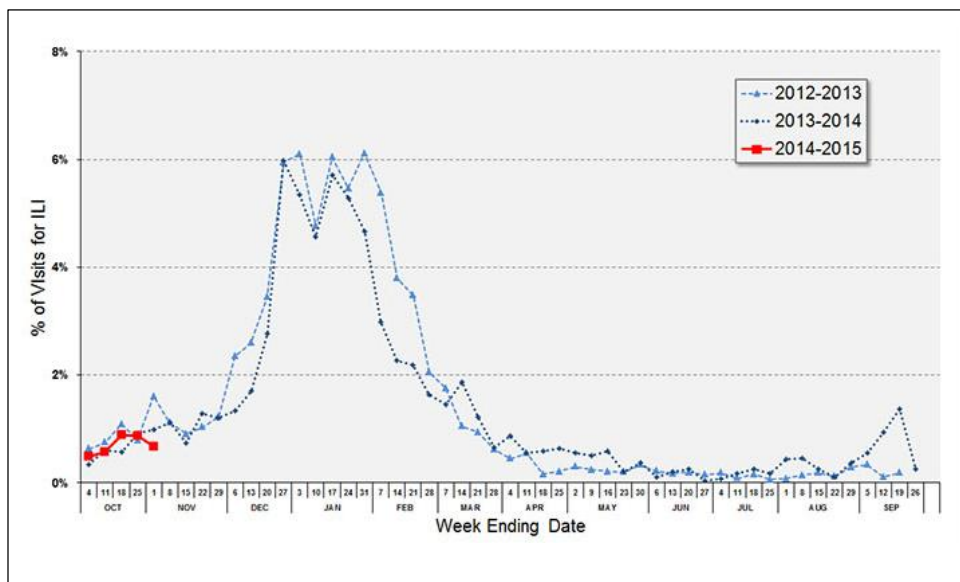


Figure 16. Analysis of influenza activity in ILINet, \*Souce CDC, <http://www.cdc.gov/flu/weekly/fluactivitysurv.htm>

of influenza activity at the local, state at national level, supporting efforts to reduce influenza associated morbidity and mortality and identification of new influenza strains, and outbreaks in the US population. ILINet could be improved by developing a real-time reporting component that would replace the weekly reports from ILINet providers and the development of tools for automated data reporting, analysis, and visualization.

#### **2.4 OTHER SYNDROMIC SURVEILLANCE SYSTEMS**

There are several other syndromic surveillance systems that have been used by federal agencies such as the CDC, and state and local health departments. These syndromic surveillance systems include RODS, EARS, BioDefend, BioStorm, BioPortal, INFERNO, and several other home-grown custom systems that states and local health departments developed for monitoring of population health ("BioSTORM: a test bed for configuring and evaluating biosurveillance methods," 2007; Dorea et al., 2014; Espino et al., 2004; Fricker et al., 2008; Naumova, O'Neil, & MacNeill, 2005; O'Connor et al., 2003; Patterson-Lomba et al., 2014; Salvadores, Alexander, Musen, & Noy, 2013). All these systems operate under the same principles of identifying specific symptoms and syndromes associated with a disease or condition and use similar detection algorithms like the ones already described.

## **CHAPTER 3: LABORATORY SURVEILLANCE SYSTEMS**

Laboratory surveillance is another form of public health surveillance. While syndromic surveillance is based on pre-diagnostic data, laboratory surveillance is based on laboratory confirmation and reporting of laboratory results (Hu et al., 2012; M. G. Johnson, Williams, Lee, & Bradley, 2014; McElwain, 2010; Niesters et al., 2013; Vogt, 1996). These laboratory data originate from clinical and public health laboratories. Laboratory surveillance is conducted for a wide range of pathogens that can induce disease such as, food and waterborne diseases, sexually transmitted and blood-borne diseases, respiratory pathogens, zoonotic diseases, etc ("Building rotavirus laboratory capacity to support the Global Rotavirus Surveillance Network," 2013; Canas et al., 2000; De Florentiis et al., 2011; Dombrowski, Buskin, Bennett, Thiede, & Golden, 2014; Hall et al., 2012; Jeremy Sueker et al., 2010; J. Lee et al., 2014; Matheny et al., 2014; Sabharwal, Braunstein, Robbins, & Shepard, 2014; Shult & Kirk, 2003). Laboratory surveillance besides monitoring population health also evaluates the impact of control measures and prevention programs against pathogens of interest. In this chapter we will provide a description of major laboratory surveillance systems currently used from federal, local, and state public health agencies.

### **3.1 NATIONAL NOTIFIABLE DISEASES SURVEILLANCE SYSTEM (NNDSS)**

The National Notifiable Diseases Surveillance System (NNDSS) is a nationwide collaboration between local, state, territorial, federal, and national public health agencies to monitor, control, and prevent the spread of nationally notifiable infectious and non-infectious conditions (N. B. Johnson et al., 2014). NNDSS was designed as a surveillance program that will collect, analyze, and share health data reported by



participating labs, but also share information related to health policies and laws, public health standards, information systems, stakeholder information, as well as processes, and resources at the local, state, and national levels (Adams et al., 2014; "Communicable diseases surveillance. Presentation of NNDSS data," 2001; Vogt, 1996). According to the CDC, NNDSS is used in order to:

- Collect, manage, share, analyze, interpret, and disseminate health-related data for state-reportable and nationally notifiable diseases and conditions.
- Develop and maintain national standards—such as consistent case definitions and electronic messaging standards.
- Monitor regional and national trends in diseases and health conditions.
- Work with other jurisdictions and partners to implement and assess prevention and control programs.
- Designate certain diseases and conditions as nationally notifiable.
- Submit data on nationally notifiable diseases to CDC.
- Maintain and publish the official national notifiable diseases statistics from 57 state, territorial, and local reporting jurisdictions in the Morbidity and Mortality Weekly Report (MMWR).

The list of reportable diseases and conditions that state and local health departments are required to report to CDC through NNDSS is shown in figure 17 (Adams et al., 2014). These conditions are classified in three different categories; conditions classified as extremely urgent and require immediate notification within four hours of laboratory confirmation by the most rapid means available, conditions classified as urgent and require notification within twenty four hours of laboratory confirmation, and conditions

classified as standard and require notification within the next normal reporting cycle.

(Centers for Disease & Prevention, 2013; Jajosky & Groseclose, 2004).

<b>IMMEDIATE, EXTREMELY URGENT - Notification within 4 hours</b>	
Call CDC EOC at 770-488-7100 within 4 hours; follow-up with electronic transmission of report by the next business day	
CONDITION	CASES REQUIRING NOTIFICATION
<b>Anthrax</b>	
Source of infection not recognized	Confirmed and probable cases
Recognized BT exposure/potential mass exposure	Confirmed and probable cases
Serious illness of naturally-occurring anthrax	Confirmed and probable cases
<b>Botulism</b>	
Foodborne (except endemic to Alaska)	All cases prior to classification
Intentional or suspected intentional release	All cases prior to classification
Infant botulism (clusters or outbreaks)	All cases prior to classification
Cases of unknown etiology/not meeting standard notification criteria	All cases prior to classification
<b>Plague</b>	
Suspected intentional release	All cases prior to classification
<b>Paralytic poliomyelitis</b>	Confirmed cases
<b>SARS - associated coronavirus</b>	All cases prior to classification
<b>Smallpox</b>	Confirmed and probable cases
<b>Tularemia</b>	
Suspected intentional release	All cases prior to classification
<b>Viral Hemorrhagic Fevers*</b>	
Suspected intentional	Confirmed and suspected cases

<b>IMMEDIATE, URGENT - Notification within 24 hours</b>	
Call CDC EOC at 770-488-7100 within 24 hours; follow-up with report in next regularly scheduled electronic transmission	
CONDITION	CASES REQUIRING NOTIFICATION
<b>Anthrax</b>	
Naturally-occurring or occupational, responding to treatment	Confirmed and probable cases
<b>Brucellosis</b>	
Multiple cases, temporally/spatially clustered	Confirmed and probable cases
<b>Diphtheria</b>	All cases prior to classification
<b>Novel influenza A virus infection, initial detections of</b>	Confirmed cases
<b>Measles</b>	Confirmed cases
<b>Poliovirus infection, nonparalytic</b>	Confirmed cases
<b>Rabies in a human</b>	Confirmed cases
<b>Rabies in an animal</b>	
Imported from outside continental US within past 60 days	Confirmed cases
<b>Rubella</b>	Confirmed cases
<b>Viral hemorrhagic fevers*</b>	
All cases other than suspected intentional	Confirmed and suspected cases
<b>Yellow Fever</b>	Confirmed and probable cases

<b>STANDARD - Notification by electronic transmission</b>	
Submit within the next normal reporting cycle (i.e., within 7 days for NNDSS conditions)	
CONDITION	CASES REQUIRING NOTIFICATION
<b>Anaplasmosis</b>	Confirmed and probable cases
<b>Arboviral disease (Calif. serogroup, EEE, Powassan, SLE, WNV, WEE)</b>	Confirmed and probable cases
<b>Babesiosis</b>	Confirmed and probable cases
<b>Botulism</b>	
Infant, sporadic cases	All cases prior to classification
Wound, sporadic cases	All cases prior to classification
<b>Brucellosis</b>	
Cases not temporally/spatially clustered	Confirmed and probable cases
<b>Cancer</b>	Confirmed cases <sup>†</sup>
<b>Carbon monoxide poisoning</b>	Confirmed and probable cases
<b>Chancroid</b>	Confirmed and probable cases
<b>Chlamydia trachomatis infection</b>	Confirmed cases
<b>Coccidioidomycosis</b>	Confirmed cases
<b>Cryptosporidiosis</b>	Confirmed and probable cases
<b>Cyclosporiasis</b>	Confirmed and probable cases

**STANDARD - Notification by electronic transmission**

Submit within the next normal reporting cycle (i.e., within 7 days for NNDSS conditions)

CONDITION	CASES REQUIRING NOTIFICATION
Dengue virus infections	Confirmed, probable and suspect cases
Ehrlichiosis	Confirmed and probable cases
<i>Escherichia coli</i> , Shiga toxin-producing (STEC)	Confirmed and probable cases
Foodborne disease outbreaks	Confirmed outbreaks <sup>‡</sup>
Giardiasis	Confirmed and probable cases
Gonorrhea	Confirmed and probable cases
<i>Haemophilus influenzae</i> , invasive disease	All cases prior to classification
Hansen's disease	Confirmed cases
Hantavirus pulmonary syndrome	Confirmed cases
Hemolytic uremic syndrome, post-diarrheal	Confirmed and probable cases
Hepatitis A, acute	Confirmed cases
Hepatitis B, acute	Confirmed cases
Hepatitis B, chronic	Confirmed and probable cases
Hepatitis B, perinatal infection	Confirmed cases
Hepatitis C, acute	Confirmed cases
Hepatitis C infection, past or present	Confirmed and probable cases
HIV infection	Confirmed cases of HIV infection; perinatally exposed infants prior to classification
Influenza-associated mortality, pediatric	Confirmed cases
Lead, exposure screening test result	All test results <sup>‡</sup>
Legionellosis	Confirmed and suspected cases
Leptospirosis	Confirmed and probable cases
Listeriosis	Confirmed cases
Lyme disease	Confirmed, probable and suspect cases
Malaria	Confirmed and suspected cases
Meningococcal disease ( <i>Neisseria meningitidis</i> )	Confirmed and probable cases
Mumps	Confirmed and probable cases
Pertussis	All cases prior to classification
Pesticide-related illness, acute	Definite, probable, possible and suspicious cases
Plague	
All cases not suspected to be intentional	All cases prior to classification
Psittacosis	Confirmed and probable cases
Q Fever	Confirmed and probable cases
Rabies in an animal	
Animal not imported within past 60 days	Confirmed cases
Rickettsiosis, Spotted Fever	Confirmed and probable cases
Rubella, congenital syndrome	Confirmed cases
Salmonellosis	Confirmed and probable cases
Shigellosis	Confirmed and probable cases
Silicosis	Confirmed cases
<i>Staphylococcus aureus</i> infection	
Vancomycin-intermediate (VISA)	Confirmed cases
Vancomycin-resistant (VRSA)	Confirmed cases
<i>Streptococcus pneumoniae</i> , invasive disease (IPD)	Confirmed cases
Streptococcal toxic shock syndrome (STSS)	Confirmed and probable cases
Syphilis	Confirmed and probable cases
Tetanus	All cases prior to classification
Toxic shock syndrome (non-Strep)	Confirmed and probable cases
Trichinellosis (Trichinosis)	All cases prior to classification
Tuberculosis	Confirmed cases
Tularemia	
All cases other than suspected intentional release	Confirmed and probable cases
Typhoid Fever	Confirmed and probable cases
Varicella	Confirmed and probable cases
<i>Vibrio cholerae</i> infection (Cholera)	Confirmed cases
Vibriosis	Confirmed and probable cases
Waterborne disease outbreaks	All outbreaks <sup>‡</sup>

\*Notifiable viral hemorrhagic fevers include those caused by Ebola or Marburg viruses, Lassa virus, Lujo virus, or new world Arenaviruses (Guanarito, Machupo, Junin, Sabia), and Crimean-Congo hemorrhagic fever

‡ Notification for all confirmed cases of cancer should be made at least annually

# Notification for lead exposure screening results should be submitted quarterly for children and twice a year for adults

‡ Outbreaks are defined by state and local health departments, all situations deemed by a local or state health department to be an outbreak are notifiable

Figure 17. List of reportable conditions, \*Source CDC and CSTE, <http://www.cste.org/>

Based on case notification reported by states to NNDSS, the CDC will analyze these data and monitor the incidence of these conditions in the US population as shown in figure 18 (N. B. Johnson et al., 2014).

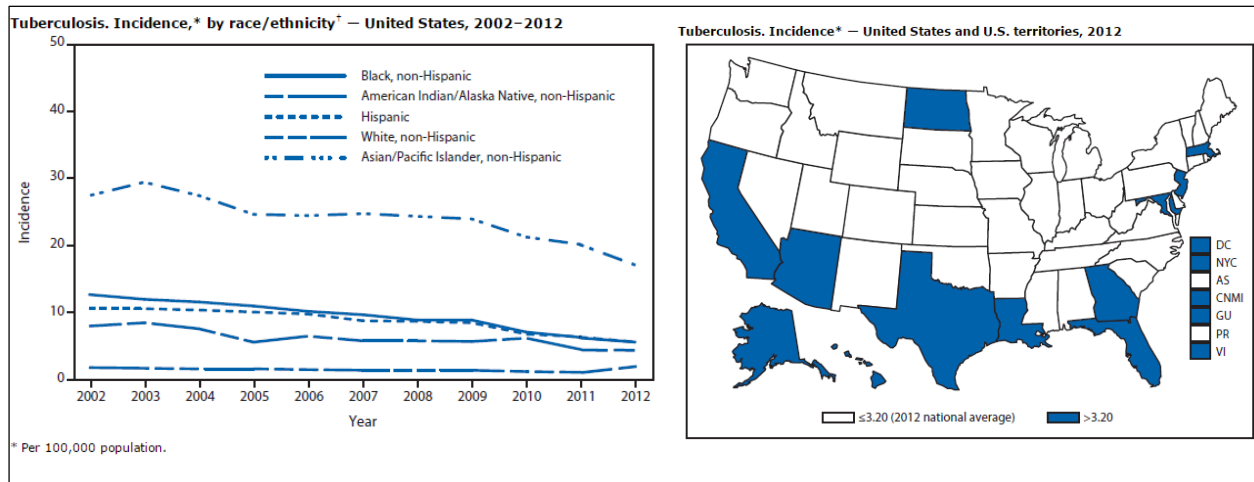


Figure 18. Analysis and visualization performed under NNDSS for MMWR (N. B. Johnson, Hayes, Brown, Hoo, & Ethier, 2014)

Based on analyzing laboratory confirmed cases, the CDC can monitor population health, evaluate the effectiveness of public health prevention and intervention programs, and propose new strategies and policies to protect and improve health of the US population. One of the key components of NNDSS is the National Electronic Disease Surveillance System (NEDSS) (National Electronic Disease Surveillance System Working, 2001; Robinson, 2014). NEDSS ensures the electronic transfer of public health surveillance data between public health departments using public health standards including Public Health Information Network (PHIN) standards and vocabulary standards such as LOINC, SNOMED, and HL7. Under NEDSS, the following requirements need to be met by public health information systems:

- Disease data entry directly through an Internet browser-based system, thereby creating a database accessible by health investigators and public health professionals.
- Electronic Laboratory Reporting (ELR) that enables labs to report cases to health departments.
- Integration of multiple health information databases into a single repository.
- Electronic messaging capabilities, enabling states to share information efficiently with CDC and other health agencies.

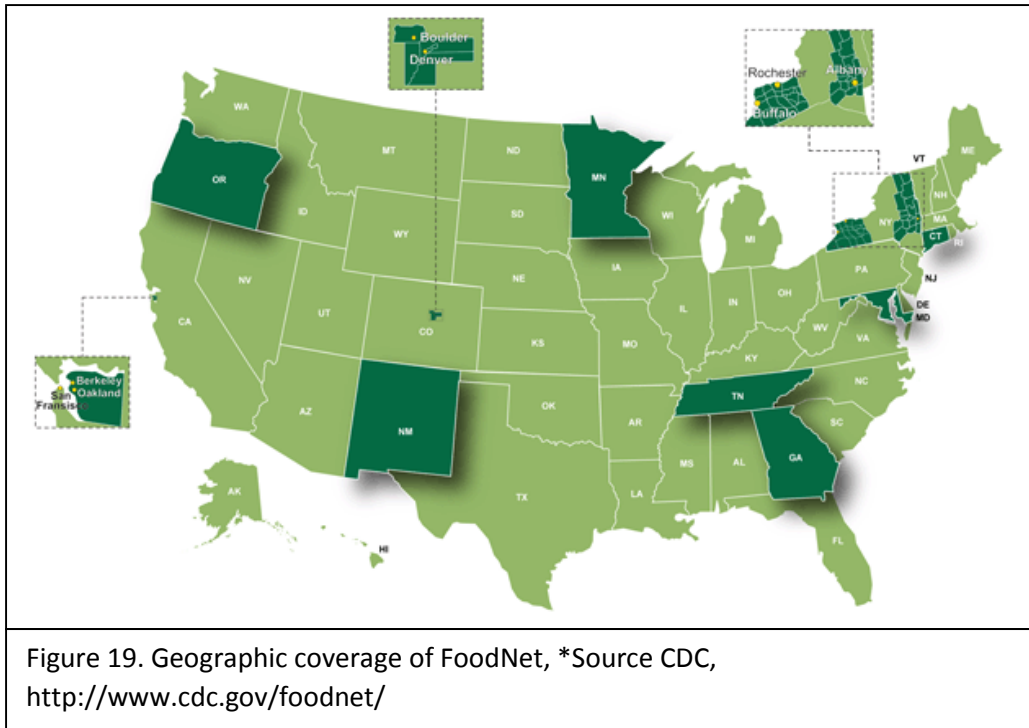
Adoption of these standards can ensure that data reported by states are shared with the CDC rapidly, securely and in a common/structured format. NNDSS could benefit from tools that would allow automated data analysis as well as dedicated systems for data visualization and geospatial analysis.

### **3.2 FOODBORNE DISEASES ACTIVE SURVEILLANCE NETWORK (FOODNET)**

The Foodborne Diseases Active Surveillance Network, or FoodNet, was developed in 1996 to track and monitor infections commonly transmitted through food including *Campylobacter*, *Cryptosporidium*, *Cyclospora*, *Listeria*, *Salmonella*, Shiga toxin-producing *Escherichia coli* (STEC) O157 and non-O157, *Shigella*, *Vibrio*, and *Yersinia* (Allos, Moore, Griffin, & Tauxe, 2004; Angulo et al., 1998; Henao et al., 2012; Manikonda et al., 2012; Scallan & Mahon, 2012; Yang, 1998). FoodNet requires diagnosis of these pathogens by laboratory testing of samples from patients. FoodNet is a collaborative effort between CDC, the Food and Drug Administration (FDA), the U.S. Department of Agriculture's Food Safety and Inspection Service (USDA-FSIS), and 10

state health departments and covers approximately 15% of the US population with more than 650 testing sites (laboratories). The 10 states enrolled in FoodNet include California, Colorado, Connecticut, Georgia, Maryland, Minnesota, New Mexico, New York, Oregon, and Tennessee as shown in figure 19.

Collected data are transmitted to CDC monthly. If the patient requires hospitalization within 7 days of the specimen



collection are also recoding with the status of the patient at hospital discharge. In addition, travel information within 7 days prior to illness is also recorded for Salmonella and STEC O157 cases (Gould, Rosenblum, Nicholas, Phan, & Jones, 2013; Henao et al., 2012; L. R. Johnson et al., 2011; Scallan & Mahon, 2012).

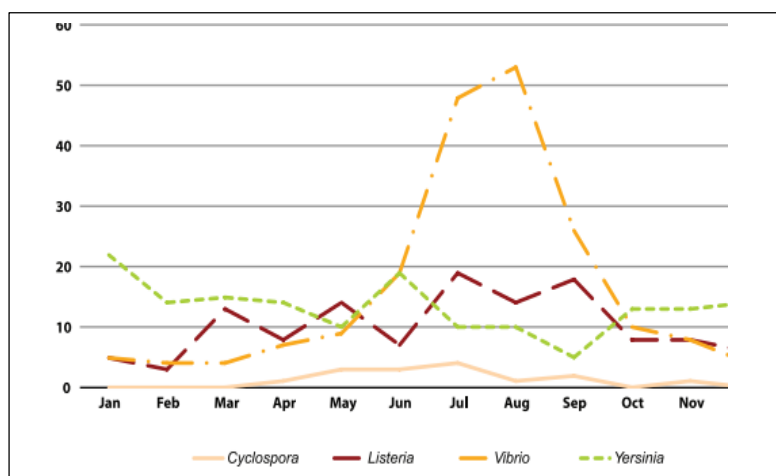


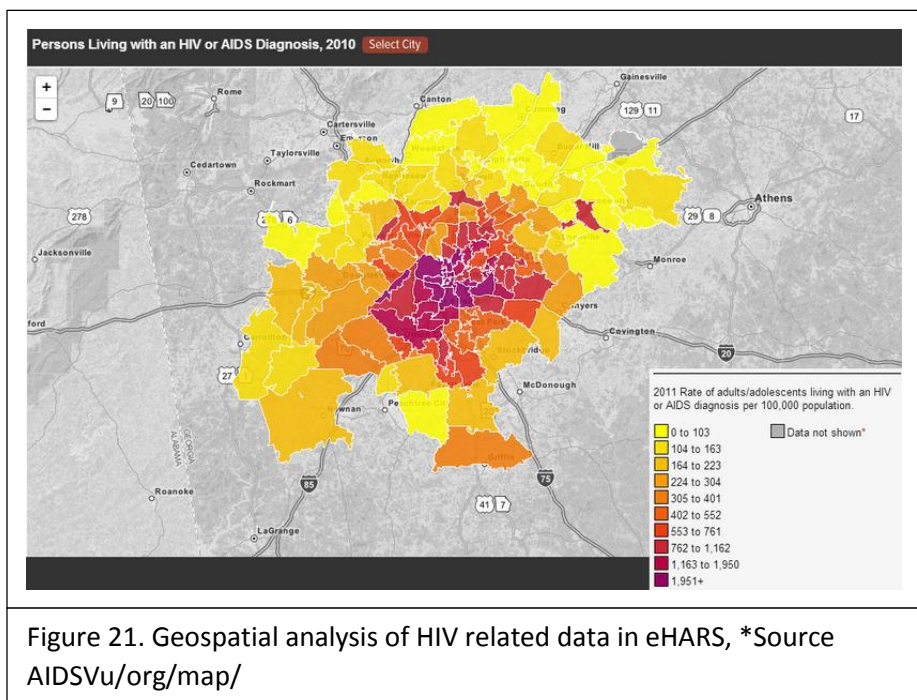
Figure 20. Analysis of foodborne diagnosed conditions in FoodNet, \*Source CDC, <http://www.cdc.gov/foodnet/>

Based on reported data, CDC monitors the activity and incidence on foodborne disease within the US population and provides annual reports related to disease incidence, affected groups of the population, seasonality, location, and other disease characteristics as show in figure 20. FoodNet not only monitors foodborne related conditions but also provides a foundation for food safety policy and prevention efforts. Overall, FoodNet could be improved by incorporating Advanced Molecular Detection (AMD) and Next Generation Sequencing (NGS) tools, more frequent reporting approaches and automated tools for data reporting, analysis, and visualization.

### **3.3 ENHANCED HIV/AIDS REPORTING SYSTEM (eHARS)**

The Enhanced HIV/AIDS Reporting System (eHARS) is the replacement surveillance system for the previous HIV/AIDS Reporting System (HARS) used by local and state health departments

to conduct HIV/AIDS case reporting, and collect data in a secure information system designed by CDC (Youmans, Tripathi, Gibson, Stephens, & Duffus, 2011).



eHARS is a relational database system that was developed in 2005 as a browser-based application to collect, manage and report HIV/AIDS case surveillance data to CDC (Mu,

Rutledge, Mills, & Paul, 2014; Youmans et al., 2011). In addition, eHARS allows for the analysis of discrete events over time, a useful feature for surveillance of chronic diseases and condition like HIV infection as shown in figure 21. eHARS supports core HIV/AIDS surveillance data activities and projects, and offers tools to support investigation of potential HIV/AIDS cases, management of current data, import and export of data, transfer of data to CDC, reporting, and analysis. eHARS collects and presents HIV/AIDS data via fields displayed in electronic documents, such as the eHARS versions of case reports, birth certificates, death certificates, and lab reports. Data collected from eHARS are used by public health practitioners and HIV planning groups for HIV/AIDS surveillance, prevalence and disease monitoring, identification of epidemiologic trends, as well as evaluation of HIV prevention programs and strategies (Mu et al., 2014; Youmans et al., 2011).

### **3.4 OTHER LABORATORY SURVEILLANCE SYSTEMS**

Laboratory surveillance is conducted for numerous pathogenes and conditions, such as the Influenza Hospitalization Surveillance Network (FluSurv-NET) for population based surveillance of laboratory confirmed influenza related hospitalizations in children and adults (Jhung et al., 2014), genomic surveillance for identifying mutations or new reassortants of different pathogenes or genetic markers in the population associated with a specific disease (Gire et al., 2014; Yuan et al., 2014), the National Respiratory and Enteric Virus Surveillance System (NREVSS) for virologic surveillance throughout the US (Rabon-Stith et al., 2013), mortality surveillance using the nationwide mortality reporting system (N. B. Johnson et al., 2014) etc. All these systems require laboratory confirmation of the disease and condition they monitor and they operate using similar standards and procedures as the one already described.



## **CHAPTER 4: WEB-BASED PUBLIC HEALTH SURVEILLANCE SYSTEMS**

Web-based public health surveillance systems are surveillance systems that were designed to collect near-real time information from internet sources and by applying data mining, machine learning, and filtering techniques, identify new cases of infectious diseases and conditions within the population (Lyon et al., 2012; Milinovich et al., 2014; Velardi et al., 2014; Wojcik et al., 2014). These systems collect information from multiple and diverse internet data sources and multiple languages such as news feeds, social media, and online reports. The amount of information flow that can be found on the internet has revolutionized how epidemic information are being collected and can offer real-time cost effective solutions to supplement existing surveillance systems or opportunities for stand-alone event-based public health surveillance systems. In multiple cases, these online data can contain individual reports and descriptions of symptoms associated with a disease or condition. By applying data mining techniques and Natural Language Processing (NLP) algorithms, public health value can be extrapolated from these data. Because these systems collect and analyze information reported online by individuals (self-reporting) related to symptoms and signs related to a disease or condition before they seek professional care and be reported by a physician or health care professional, they precede syndromic surveillance systems (Milinovich et al., 2014). Analyzing these various information from diverse internet sources and using internet based tools, initial evidence and signs of an outbreak can be detected. In addition, these systems can be proven valuable in areas that are not currently performing any type of public health or the capabilities of these systems are limited. Although these systems have been associated with early disease detection and

outbreaks in previous cases (MERC-CoV, Ebola, etc.) and served as sources to enhance existing syndromic surveillance systems, they have also been associated with high levels of false positive reporting and other issues associated with detection sensitivity, signal detection, and background noise. One of the most well-known cases of false reporting and difficulties with accurate signal detection and separation from noise is google flu trends (Santillana, Zhang, Althouse, & Ayers, 2014). Google flu trends, owned and operated by Google Inc. claimed that just by collecting search terms related to influenza disease, they could accurately predict influenza activity in the population. Google flu trends model is based on the relationship between how many people search for influenza related topics and actually positive influenza cases. This is a linear model that compute the log-odds of physician visits associated with influenza like illness (ILI) and the log-odds of ILI-related search query (Fearhead, Giagos, & Sherlock, 2014). Despite successful initial results published in high impact medical journals that predicted influenza activity in the US (Carneiro & Mylonakis, 2009), further analysis demonstrated that google flu trends made inaccurate forecasts for at least 100 of 108 weeks and overestimating influenza cases for as much as 50% in many cases (Lazer, Kennedy, King, & Vespignani, 2014). The reason behind this overestimation and incorrect forecasting was located in the amount of noise behind the big data analyzed by google (Lazer et al., 2014; Santillana et al., 2014). Developing algorithms that can accurately calculate the signal-to-noise ratio in these systems is crucial for building accurate web-based public health predictive models. To reduce the noise, these systems require continuous analysis of web-based data and validation with an alternative high quality set of data. Furthermore, these systems require constant

modification of its predictive and machine learning algorithms based on the changes recorded on social patterns from the various online sources. Based on these changes, re-training and recalibrating these models and adapting them based on all new parameters is required to maintain the high accuracy of the system.

Despite these challenges, these web data mining and crowd-sourced public health tracking systems can provide the real-time signal indicative of a disease or condition in the population that syndromic or laboratory surveillance systems lack. In this chapter we will provide a description of major web-based surveillance systems with significant impact and results in early disease detection. .

#### **4.1 HEALTHMAP**

HealthMap is an automated system established in 2006 that collects and display information about new outbreaks in humans and animals according to geographic location, time, and infectious agent (Lyon et al., 2012). HealthMap has been funded in part by Google.org, and has been collaborating with the U.S. Department of Health and

<b>HealthMap Data Sources</b>	
Google News	WHO alerts
RSS feeds	Weather reports
ProMED-mail	News articles
Eurosurveillance	Twitter feeds/Social media
Community news reports	SOSO
User eyewitness reports	Baidu

Figure 22. HealthMap primary data sources

Human Services to map seasonal influenza and H1N1 in the U.S. The system integrates outbreak data from multiple electronic sources, including online news wires (e.g., Google News), Really Simple Syndication (RSS) feeds, expert-curated accounts (e.g., ProMED-

mail, a global electronic mailing list that receives and summarizes reports on disease outbreaks), multinational surveillance reports (e.g., Eurosurveillance), and validated official alerts (e.g., from WHO). Overall, HealthMap collects information from Baidu, EuroSurveillance, Google, Community News Reports, OIE, ProMED-mail, SOSO, User Eyewitness Reports, WDIN and WHO as shown in figure 22 and scans for articles in multiple languages.

It also has a mapping system that allows users to view reports and apply a number of filters. Users can also comment on articles and rank them for significance (Brownstein & Freifeld, 2007; Brownstein, Freifeld, Reis, & Mandl, 2008; Freifeld et al., 2008; Lyon et al., 2012; Morse, 2012).

As a result, HealthMap collects data in real time from more than 20 000 websites daily and processes an average of 133.5 disease alerts/day with approximately 50%

categorized as breaking news (65.3 reports per day) that can be plotted on an interactive



Figure 23. Visualization of results in HealthMaps' main user interface, \*Source HealthMap, <http://healthmap.org/en/>

map based on the Google Maps API as seen in figure 23 (Brownstein & Freifeld, 2007; Freifeld et al., 2008; Lyon et al., 2012).

Surveillance is conducted in several languages, including Arabic, Chinese, English, French, Portuguese, Russian and Spanish and there are currently efforts to expand the system to collect data from other languages. The system receives 1,000–10,000 visits/day from around the world with the most frequent visitors originating from government-related domains, including WHO, CDC, European Centre for Disease Prevention and Control, and other national, state, and local bodies worldwide. HealthMap also provides advanced filtering capabilities. After the first categorization step into locations and diseases, a second round of category tags is applied to the articles to improve filtering. The primary tags include 1) breaking news (e.g., a newly discovered outbreak); 2) warning (initial concerns of disease emergence, e.g., in a natural disaster area); 3) follow-up (reference to a past outbreak); 4) background/context (information on disease context, e.g., preparedness planning); and 5) not disease-related (information not relating to any disease). Duplicate reports are also removed by calculating a similarity score based on text and category matching. The tagged data gets analyzed using NLP methodologies, detection algorithms, and predictive analysis tools for disease and condition detection among the population. Finally, in addition to providing mapped content, each alert is linked to a related information window with details on reports of similar content as well as recent reports concerning either the same disease or location and links for further research (e.g., WHO, CDC, and PubMed). Finally, it can represent the locations of events with a coloured marker depending on disease severity and event type.

One of the strongest advantages of this system in collaboration with other international organizations including CrisisMappers and Humanity Road, is the integration of social

media reporting during disasters (for example social media integration during the Haiti cholera outbreak) that allows real-time data collection and detailed health information in a crisis setting. As a result, the web resource ([healthmap.org/haiti](http://healthmap.org/haiti)) mapped informal outbreak reports from news and social media as well as key information on the availability of healthcare and clean water facilities.

HealthMap was the first surveillance system that detected reports about hemorrhagic fever cases in southeastern Guinea in March 14 2014, 9 days before the World Health Organization (WHO) formally announced the Ebola outbreak, the largest Ebola epidemic ever reported in history.

#### **4.2 ProMED-mail**

Program for Monitoring Emerging Diseases (ProMED) is an early warning and disease-reporting system that was established in 1994 with the support of the Federation of American Scientists and SatelLife (Pollack, Pringle, Madoff, & Memish, 2013). Since October 1999, ProMED-mail has operated as an official program of the International Society for Infectious Diseases, a nonprofit professional organization with 20,000 members worldwide (Stewart & Denecke, 2010). ProMED monitors outbreaks of

<b>ProMED-mail Data Sources</b>	
Media reports	Official reports
Online summaries	Local observers
Subscribed users	News articles

Figure 24. ProMED-mail primary sources of information and data collection

infectious diseases that can affect humans, animals or plants but also acute exposures to toxins that affect human health, including those in animals and in plants grown for food or animal feed (Cowen et al., 2006). The system's sources of information include

media reports, official reports, online summaries, local observers, and others. Reports are often contributed by ProMED-mail subscribers as shown in figure 24. ProMED-mail has been supported by various organizations including Google.org, the Gates Foundation, the Rockefeller Foundation, the Oracle Corporation, the Nuclear Threat Initiative, and others and closely collaborates with HealthMap.

ProMED-mail collects data in real time that have been submitted to health agencies and international organizations, media sources or reports submitted directly from subscribers. These reports are initially being screened by a moderator for their validity and can either reject them or send them further to subject moderators that will categorize and analyze them accordingly. There are 12 different categories that include 4 for veterinary and zoonotic diseases, 2 for viral diseases, and 1 each for bacterial diseases, parasitic diseases, plant diseases, epidemiology and surveillance, and medical entomology. Upon analysis of the data, a report will be generated with the description of the event as seen

in figure 25 (Antohi et al., 2007).

Reports generated by the system are distributed by email to direct

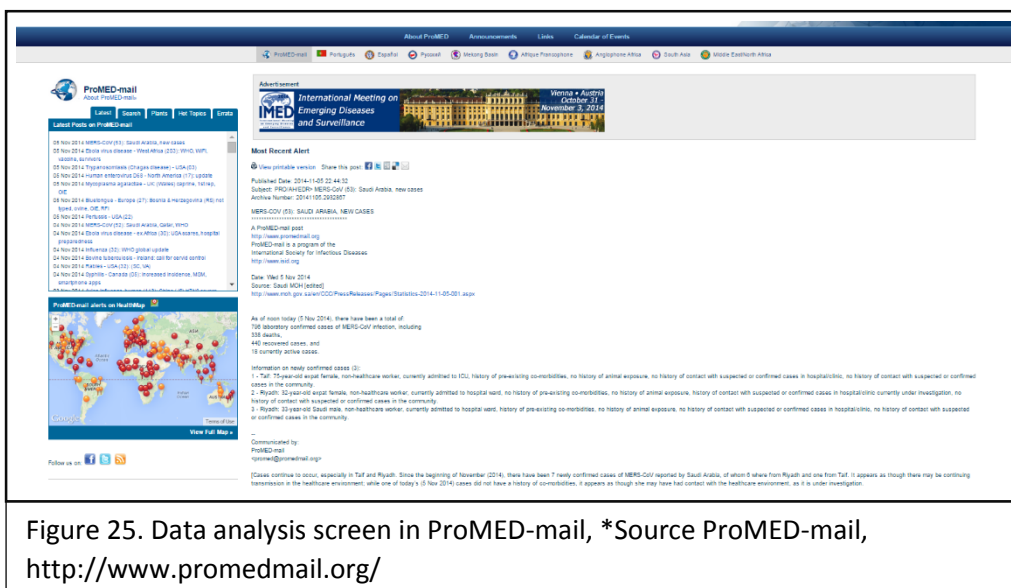


Figure 25. Data analysis screen in ProMED-mail, \*Source ProMED-mail, <http://www.promedmail.org/>

subscribers and posted immediately on the ProMED-mail web site. ProMED-mail currently reaches over 60,000 subscribers in at least 185 countries and the geographic

visualization of the data is accomplished through a partnership with Healthmap.org (Antohi et al., 2007; Cowen et al., 2006; Freifeld et al., 2008; Madoff, 2004; Stewart & Denecke, 2010).

ProMED-mail was the first surveillance system to report cases of individuals getting sick with symptoms associated with an unknown pathogen similar to the ones reported during the severe acute respiratory syndrome coronavirus (SARS) outbreak. Later on it was identified that these cases were induced by a new virus strain, the Middle East respiratory syndrome coronavirus (MERS-CoV) (Pollack et al., 2013).

### **4.3 BioCaster**

BioCaster is an automated early warning system that was established in 2006. The goal of the system is monitoring of biological events that affect humans, animals and plants as a result of chemical or radio nuclear event or after natural disasters such as earthquakes, typhoons, floods etc (Collier et al., 2008; Lyon et al., 2012). The system is supported and operates from the Japanese National Institute of Informatics (NII) in Tokyo and the Japan Science and Technology Agency’s PRESTO fund. BioCaster collects data in real time from a variety of sources such as Google News, World Health Organization (WHO), MeltWater News, ProMED-mail, European Media Monitor Alerts (MedISys) as shown in figure 26.

<b>BioCaster Data Sources</b>	
Google News	WHO
MeltWater News	ProMED-mail
European Media Monitor Alerts (MedISys)	RSS feeds

Figure 26. BioCaster primary sources of information and data collection



The system continuously analyzes documents reported from over 1700 RSS feeds daily in 10 different languages including Arabic, Chinese, English, French, Japanese, Korean, Portuguese, Russian, Spanish, Thai and Vietnamese and classifies them for topical relevance and plots onto a Google Map using geocoded information.

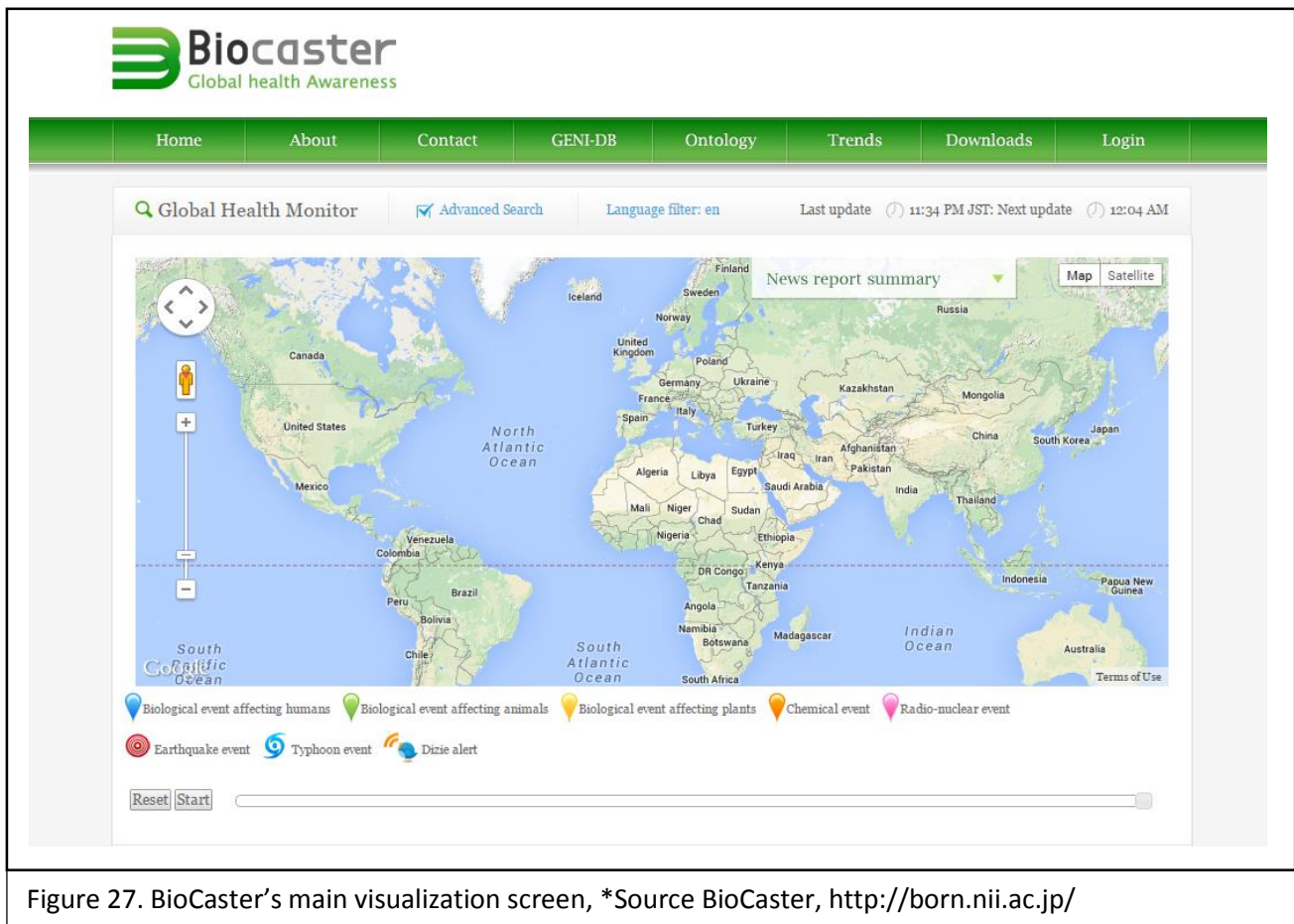


Figure 27. BioCaster’s main visualization screen, \*Source BioCaster, <http://born.nii.ac.jp/>

BioCaster contains a web/database server and a backend cluster computer equipped with a variety of text mining algorithms which continuously scan hundreds of RSS newsfeeds from local and national news providers. Since the text mining system has a detailed knowledge about the important concepts such as diseases, pathogens, symptoms, people, places, and drugs. This allow to semantically index relevant parts of news articles, enabling users to have quicker and highly precise access to information

(Collier et al., 2008; Lyon et al., 2012). The reported data comes from annotated text collections, gazetteer lists of nomenclature and the BioCaster ontology, all of which are currently under development. Users can access the results by accessing the BioCaster site as shown in figure 27, where they are presented with a “live” map of events occurred over the last 30 days. The system provides filtering capabilities based on date, language, geographical location, and type of disaster and other parameters for improved sensitivity and specificity of the system in disease detection.

#### **4.4 OTHER WEB-BASED SURVEILLANCE SYSTEMS**

Currently, there are several web-based biosecurity intelligence systems that collect information related to disease outbreak. All these systems have as a main goal to gather and analyze information relevant to public health and provide an early warning during an outbreak. Some other web-based surveillance systems besides the ones described include EpiSPIDER, MappyHealth, MedISys, the Global Public Health Intelligence Network (GPHIN), USHAHIDI, and others (Lyon et al., 2012; Ting, Tsang, Ip, & Ho, 2011).

Although these systems operate under similar principals of monitoring, collecting and reporting of data, there are significant differences related to the sources and languages these systems use to collect the information, the geographic regions it covers, but also differences in the NLP algorithms, machine learning techniques, data filtering approaches, and the methods to analyze data. The plethora of systems operating in this space points out to the possibilities offered for public health surveillance by collecting and analyzing social media and other self-reporting information in real-time.

## **CHAPTER 5: THE FUTURE OF PUBLIC HEALTH SURVEILLANCE SYSTEMS**

Public health surveillance systems have transformed the way we conduct public health. Since their inception, public health surveillance systems have evolved from tools for warning of possible bioterrorism attacks, to tools for early detection not only of infectious diseases, but health conditions affecting the population in general. As a result, public health surveillance systems have become the perfect tools for monitoring infectious diseases but also not infectious conditions such as injuries, chronic conditions, toxic exposures, drug use, environmental and occupational exposures, birth defects, mental illness, etc. Thus, they can provide an overall assessment of population health, situational awareness, rapid decision making support, and public health policy evaluation and support.

Although these systems were built using the best tools, technologies, and system architecture standards that were available during their design, several technological advancements in the areas of predictive analytics, Electronic Health Records (EHRs), as well as the development of NoSQL databases can offer new and highly effective enhancement tools to existing systems or alternative solutions for re-designing new public health surveillance systems that would incorporate these new technologies

### **5.1 PREDICTIVE ANALYTICS IN PUBLIC HEALTH**

Predictive analytics is a new area of interest in public health surveillance that combines new technologies such as machine learning (Marella, Sparnon, & Finley, 2014; Z. Wang et al., 2012; Worden & Manson, 2007), predictive modeling (Farran, Channanath, Behbehani, & Thanaraj, 2013; Mathias et al., 2013; Tabak, Sun, Nunez, & Johannes,

2014), and data mining (Lucas, 2004; Partington, Papakroni, & Menzies, 2014; Velardi et al., 2014), in order to analyze current and historical public health data and to make predictions about future health trends, enhance automated data analysis, and improve sensitivity when analyzing large data sets.

Machine learning is a computer science discipline that focuses on developing algorithms that will allow the automated process of learning from analyzed data (Boxwala, Kim, Grillo, & Ohno-Machado, 2011; Fidahussein & Vreeman, 2014; Khondoker, Dobson, Skirrow, Simmons, & Stahl, 2013; Marella et al., 2014). This approach can find great application in automatic detection of public health events by analyzing massive sets of data that contain public health related information. Once the system “learns” the parameters and the distinction between normal/baseline and abnormal conditions based on developing and optimizing these learning algorithms, it will be capable of automatic monitoring, detection, and reporting of public health events of interest. There are several machine learning techniques currently used in public health systems, such as classifiers, clustering, Bayesian statistic, and genetic algorithms (Lucas, 2004).

Predictive modeling is a mathematical method by which different statistical models are developed to try to predict the probability of an outcome based on available data. In the case of public health these outcomes may be disease spread, morbidity and mortality, impact on the population, identification of high risk groups, etc. (Andersson, Faverjon, Vial, Legrand, & Leblond, 2014; Burr et al., 2006; Miller et al., 2007; Perry, Korenberg, Hall, & Moore, 2011; Tabak et al., 2014; Valencia-Mendoza & Bertozzi, 2008).

Predictive models collect data and look for mathematical relationships between

dependent variables and various independent variables and measuring the probability of these relationships to occur in the future. In addition, since these relationships are never perfect in practice, certain degree of improbability is added to the model. There are several different predictive modeling techniques that can be used such as, lineal regression, K-means clustering, traditional decision trees, neural networks, random forests, etc. All these predictive modeling techniques use different algorithms and different levels of relationships between the data.

Data mining is the process of exploring large amounts of data/Big Data in search of identifying consistent patterns and relationships between data elements (Cubillas, Ramos, Feito, & Urena, 2014; Cunningham et al., 2014; Gotz, Wang, & Perer, 2014; Partington et al., 2014). Once a pattern is identified, it gets validated by applying it in a new set of data. The main goals behind data mining are prediction and predictive data mining, but also finding value behind non-traditional sources/types of data. Data mining is a computationally intense process that requires several prior steps related to data preparation, data reduction, and data analysis.

Because of the value these new technologies bring in automated data analysis and reporting, there is great interest in adopting/implementing them in healthcare and public health surveillance. Currently, these techniques are being used to some degree from web-based surveillance systems that use machine learning, NLP, data mining, and predictive models to identify and predict potential threats circulating within the general population.

## **5.2 ELECTRONIC HEALTH RECORDS (EHRs)**

EHRs are digital versions of patient's records containing the entire medical history as well as other personal identifiable (sensitive) information (Goedert, 2008; Hayrinen & Mykkanen, 2011; Lanham, Leykum, & McDaniel, 2012; Mold et al., 2012; Sittig, Singh, & Longhurst, 2013; Spyropoulos et al., 2014). EHRs contain a patient's full medical history, medication, insurance plans, treatment plans, diagnoses, imaging and laboratory results, immunization dates, provider information, etc. Adoption of EHRs has been an important initiative and adoption of EHR systems has been incentivized and facilitated by the Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States and offers incentives to providers that adopt and use EHRs in clinical settings (Barnes, 2011; Golder, 2010; Joseph, Snow, Furukawa, Posnack, & Chaffee, 2014; Mehta, 2010; Terry, 2010; T. Wang, Wang, & Biedermann, 2013). The main goal of this effort is to modernize the health system by promoting and expanding the adoption of health information technology and reduce medical costs. EHRs adopt existing standards about message transport mechanisms (HL7), ICD-9 diagnostic codes (transition to ICD-10), and laboratory reporting using LOINC and SNOMED codes. Since EHRs contain highly structured health related data, they can ensure interoperability with existing public health surveillance systems and further enhance their functionality. This can be achieved by rapidly reporting chief complaints as captured by ICD-9 and ICD-10 codes, laboratory orders, and laboratory results. In addition, the high quality and completeness of EHRs can reduce the time required for data validation and data curation and support rapid analysis. All these advantages can improve system functionality, reduce reporting lags, and decrease operating costs.

### **5.3 NOT ONLY STRUCTURED QUERY LANGUAGE (NoSQL) DATABASES**

NoSQL are modern database/data model systems for document stores, key value stores, XML databases, graph databases, column stores, and object stores. They assume that data storage does not require fixed table schemas and do not follow the relational database management systems traditional SQL databases require (K. K. Lee, Tang, & Choi, 2013; Ningthoujam et al., 2014). Since NoSQL systems do not have to follow these requirements, they have a simpler design, better data management system, dynamic schemas (or schema-less), are horizontally scalable, and have an open architecture. As a result, NoSQL databases can store and handle/analyze large volumes of not only structured but also unstructured data (data that do not fit under fixed table schemas). In addition, since NoSQL architecture allows for scaling and distributed use across a large number of servers, they can support parallel processing of massive volumes of data in cloud instances, virtual machines, and servers, thus improving performance of data analysis when compared to SQL based systems. Several NoSQL solutions have been developed for different database types and purposes such as MongoDB, Cassandra, and Apache Hadoop (Dong et al., 2013; Ningthoujam et al., 2014). Because of these benefits, NoSQL databases have been gaining in popularity in healthcare applications over the traditional SQL databases

Implementing these technologies in existing public health surveillance systems or designing new systems based on these architectures would be the next major step for public health surveillance. These tools have the potential to improve data reporting times, eliminate lack of interoperability between systems due to data format limitations, introduce predictive modeling tools for disease spread within the population as

integrated parts of surveillance systems, and support automated data analysis. The maturity level, reliability, and potential these technologies offer and the possibility of implementing them in existing system architectures or using them as the base for designing new unified systems, could advance public health surveillance systems into tools for near real-time monitoring of population health and disease outbreaks, with enhanced situational awareness and rapid decision making support and response capabilities.

#### **5.4 DESIGNING THE IDEAL PUBLIC HEALTH SURVEILLANCE SYSTEM**

Throughout the description of these public health surveillance systems for syndromic, laboratory, and web-based surveillance, several limitation can be identified. A summary of these limitations is show on figure 28.

<b>LIMITATIONS OF PUBLIC HEALTH SURVEILLANCE SYSTEMS</b>	
<b>Interoperability</b>	Lack of data exchange, data sharing, data interpretation, synchronous operation, or communication between different systems
<b>Data Sources</b>	Use of different and diverse data sources between different systems
<b>Data Collection</b>	Use of different approaches, mechanisms, and methodologies, for scanning and collecting data from sources
<b>Data Reporting</b>	Lag in data reporting and lack of visualization tools for certain surveillance systems
<b>Data Analysis</b>	Lack of automated data analysis or tools to support data analysis during data collection. Lag in data analysis
<b>Data Quality</b>	Lack of validation mechanisms for data completeness, and data duplication
<b>Signal Detection</b>	Different signal detection algorithms with different levels of sensitivity and specificity for each system
<b>Automation</b>	Lack of tools for automated data ingestion, data analysis, and data reporting
<b>Standardization</b>	Use of different standards for data collection, data/message transport, and data reporting
<b>Data Ownership</b>	Policies, regulations, and legal frameworks regarding data ownership rights
<b>Data Security</b>	Policies and regulations related to data storage, data transport, and data sharing of data containing PII and PHI

Figure 28. Limitation and challenges of public health surveillance systems



Currently, public health surveillance systems that provide nationwide coverage operate under a model that focuses on distributing all surveillance data and efforts in a centralized location within the CDC. Under this model, public health providers that collect public health related data such as chief-complains, disease symptoms and syndromes, laboratory orders, laboratory sample/specimen collection, laboratory results, etc. are required to share with public health professional from the federal government. Based on public health policies and regulations related to data ownership rights, any sharing of public health data outside the state level requires the approval and release of these data at the local and state level. As a result, all public health data require to be reviewed and approved by local and state health officials before being shared with CDC and other federal stakeholders.

In addition, due to safety concerns and regulations about PII and PHI as defined under HIPAA, all PII and PHI are removed from this data. As a results, further linking this data to laboratory results and laboratory based surveillance becomes impossible. As a result, syndromic surveillance is disconnected from laboratory surveillance.

Once the data gets released from the state and local health departments to CDC, further analysis of these data including establishing a functional signal background, application of detection algorithms, and identification of potential diseases and conditions within the population can be achieved. In addition, normalization of data collected from various sources and conversion into a common format that can support data ingestion and analysis from this centralized system occurs during this stage, introducing further delays in abnormal signal detection within the population. Furthermore, CDC and federal

stakeholder cannot address potential issues with the data directly with the health care providers since the involvement of local and states officials is required.

All these limitations result in serious lags in data transport, data analysis, and data reporting, and overall delays in early disease detection, laboratory confirmation, and decision making support. This linear condition model of public health surveillance between health care providers, local and state health departments, and CDC and federal stakeholders is shown in figure 29.

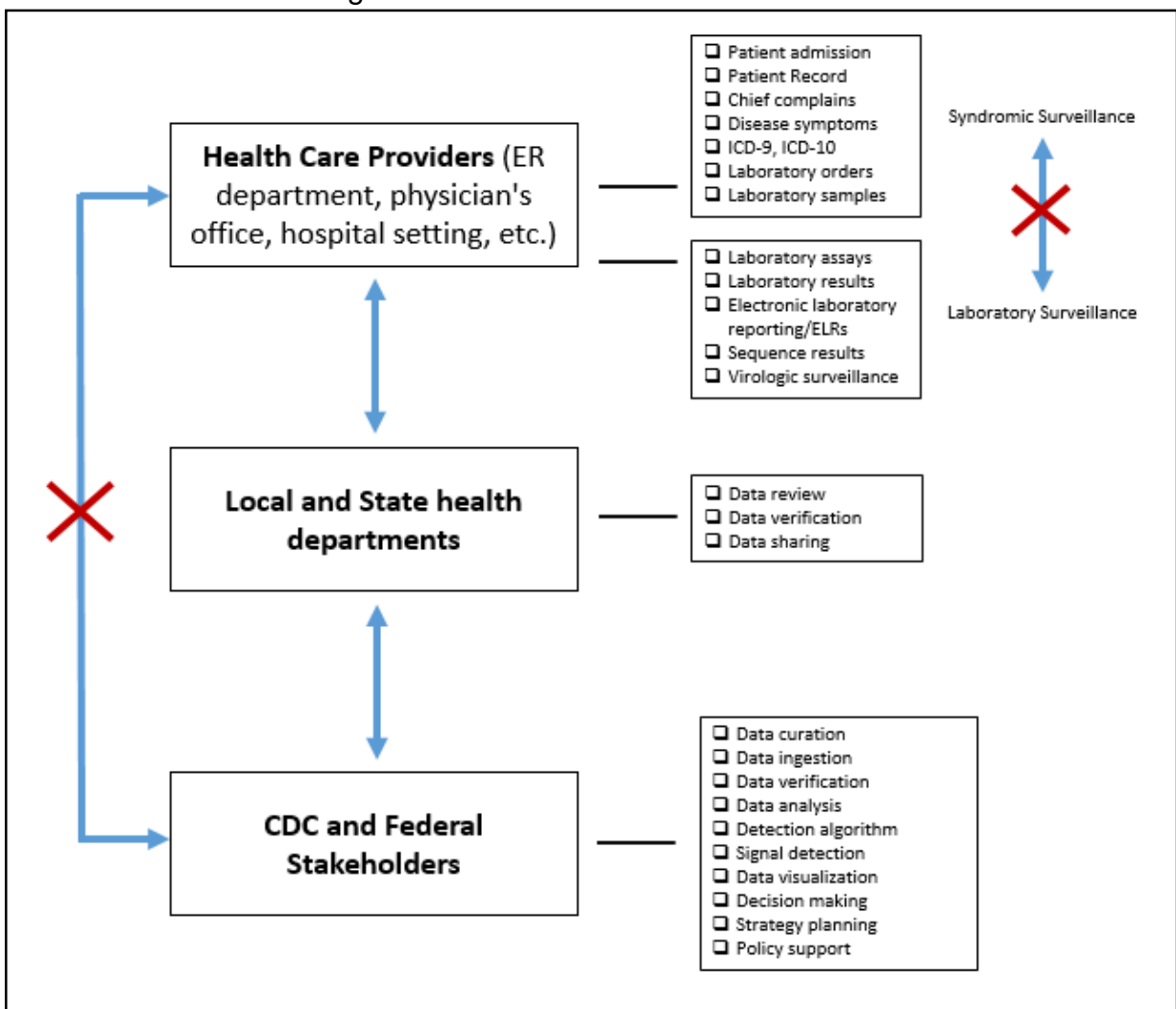


Figure 29. Current model of public health surveillance

Although resolving several of these limitations can be a major challenge, a possible solution could be achieved by decentralizing public health surveillance and crowdsourcing it to the state, local, and hospital setting level as a more agile and effective approach for conducting public health surveillance. This would allow for easier and cost effective integration of automated solutions for data collection and analysis in the existing systems currently used at the local level. Suspected cases could be “flagged” as potential threats and only these cases would be reported at a higher level such as the CDC for further investigation. Under this model, data collection, analysis, and reporting would be performed at the local level. Detection of signals that could indicate a potential disease or condition in the population would be directly reported to both local and state health departments as well as CDC and federal stakeholders. A monthly data dump would provide access to un-identified PII and PHI data to federal stakeholders for further analysis, longitudinal studies, and archiving purposes. This model could overcome limitations associated with data ownership and data reporting while it would still provide situational awareness and rapid decision making support to federal stakeholders.

There are several advantages that can be identified with this approach. Since disease trends and conditions can be significantly different among different states, jurisdictions, or different groups in the local population depending on local policies and prevention strategies, socio-economic factors, but also geographic location, demographics of the population, and disease seasonality, the background baseline for different diseases could be different between different states. For example, the rates of cancer, HIV, or influenza can be significantly different depending on environmental factors, prevention and educational programs and strategies, or vaccination coverage. Defining the baseline

based on local parameters or developing detection algorithms based on these parameters can be a much more effective approach to improve sensitivity and specificity of public health surveillance systems but also allow for immediate and more organized response to population health and community needs when compared to national surveillance. Furthermore, this approach would allow for better understanding of local population health needs and shaping of local health policies and intervention strategies that could directly target high priority problems. In addition, since this approach is performed at a local level, it would allow the connection of both syndromic and laboratory surveillance under the same system. This can further ensure the validity of syndromic surveillance models by constantly comparing them against high quality laboratory confirmed data. In addition, this model would allow for more direct and immediate engagement of medical professionals and subject matter experts with issues related to population health, and evaluating potential abnormal signals detected in the population.

From a system implementation stand point, although previously described tools such as predictive analytics, EHR adoption, NoSQL database architecture, cloud computing and analytics represent highly effective solutions that can improve system automation throughout the entire process of data collection, validation, analysis, and reporting, implementation of these solutions under a unified system with multiple parameters and dependencies from existing systems can be challenging. Implementing these tools in smaller scale systems that can be tailored according to each system's parameters would be a more compatible and cost-effective approach than a "one size fits all" solution.

Finally, this model would allow for updating single system components without requiring major system re-designs, or compromising the functionality of the entire system.

Transferring surveillance actions and responsibilities to local level could be supported through The Office of the National Coordinator for Health Information Technology (ONC) initiatives, such as the HITECH act for IT adoption, the Meaningful Use initiative under The American Recover and Reinvestment Act, or the Health Information Exchange (HIE) initiative. Through these initiatives, incentives could be offered to local health departments and clinics to overtake this important function of performing real-time processing and analysis of health related data and support public health surveillance efforts.

This multilateral/bidirectional model of public health surveillance between health care providers, local and state health departments, and CDC and federal stakeholder is shown in figure 30.

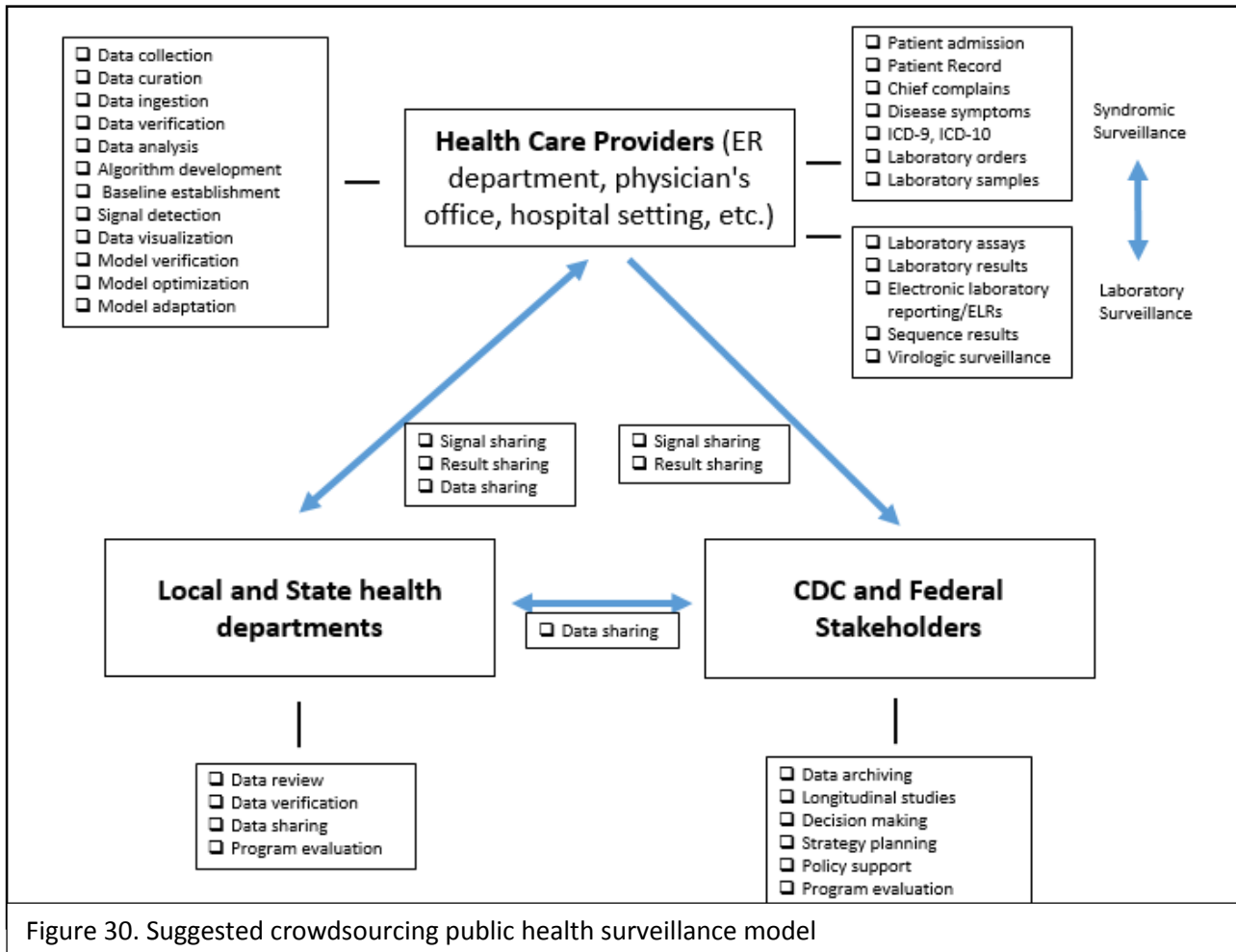


Figure 30. Suggested crowdsourcing public health surveillance model

Public health surveillance systems play a crucial role in protecting the population from potential threats and conditions. Improving these systems based on the knowledge and experience that we have gained so far, integrating new technologies and tools that can further enhance and improve their functionality, but also supporting policy changes related to health data ownership and sharing would be the next necessary steps that will transform these systems to valuable tools for rapid disease detection, situational awareness, and decision making support.

## **REFERENCES**

- The National Academies Press. (2007) *Global Infectious Disease Surveillance and Detection: Assessing the Challenges-Finding Solutions, Workshop Summary*. Washington (DC).
- Adams, D. A., Jajosky, R. A., Ajani, U., Kriseman, J., Sharp, P., Onwen, D. H., . . . Abellera, J. P. (2014). Summary of notifiable diseases--United States, 2012. *MMWR Morb Mortal Wkly Rep*, 61(53), 1-121.
- Addressing emerging infectious disease threats: a prevention strategy for the United States. Executive summary. (1994). *MMWR Recomm Rep*, 43(RR-5), 1-18.
- Allos, B. M., Moore, M. R., Griffin, P. M., & Tauxe, R. V. (2004). Surveillance for sporadic foodborne disease in the 21st century: the FoodNet perspective. *Clin Infect Dis*, 38 Suppl 3, S115-120. doi: 10.1086/381577
- Alwan, A., Maclean, D. R., Riley, L. M., d'Espaignet, E. T., Mathers, C. D., Stevens, G. A., & Bettcher, D. (2010). Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *Lancet*, 376(9755), 1861-1868. doi: 10.1016/S0140-6736(10)61853-3
- Andersson, M. G., Faverjon, C., Vial, F., Legrand, L., & Leblond, A. (2014). Using bayes' rule to define the value of evidence from syndromic surveillance. *PLoS One*, 9(11), e111335. doi: 10.1371/journal.pone.0111335
- Andersson, T., Bjelkmar, P., Hulth, A., Lindh, J., Stenmark, S., & Widerstrom, M. (2014). Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries

and over-the-counter pharmacy sales. *Epidemiol Infect*, 142(2), 303-313. doi: 10.1017/S0950268813001088

Angulo, F. J., Voetsch, A. C., Vugia, D., Hadler, J. L., Farley, M., Hedberg, C., . . . Swerdlow, D. L. (1998). Determining the burden of human illness from food borne diseases. CDC's emerging infectious disease program Food Borne Diseases Active Surveillance Network (FoodNet). *Vet Clin North Am Food Anim Pract*, 14(1), 165-172.

Antohi, R., Ogescu, C., Stefan, L., Raureanu, M., Onofriescu, M., & Toma, M. (2007). Management of the electronic patient records in the web based platform for diagnosis and medical decision for optimization in healthcare-PROMED. *Stud Health Technol Inform*, 127, 98-107.

Barnes, P. E. (2011). Think HITECH-HIPAA changes don't apply to you? Think again! *J Miss State Med Assoc*, 52(12), 382-385.

Benoit, S. R., McDonald, L. C., English, R., & Tokars, J. I. (2011). Automated surveillance of *Clostridium difficile* infections using BioSense. *Infect Control Hosp Epidemiol*, 32(1), 26-33. doi: 10.1086/657633

Berkelman, R. L., & Buehler, J. W. (1990). Public health surveillance of non-infectious chronic diseases: the potential to detect rapid changes in disease burden. *Int J Epidemiol*, 19(3), 628-635.

Betancourt, J. A., Hakre, S., Polyak, C. S., & Pavlin, J. A. (2007). Evaluation of ICD-9 codes for syndromic surveillance in the electronic surveillance system for the early notification of community-based epidemics. *Mil Med*, 172(4), 346-352.



- BioSTORM: a test bed for configuring and evaluating biosurveillance methods. (2007).  
*AMIA Annu Symp Proc*, 1139.
- Boxwala, A. A., Kim, J., Grillo, J. M., & Ohno-Machado, L. (2011). Using statistical and machine learning to help institutions detect suspicious access to electronic health records. *J Am Med Inform Assoc*, 18(4), 498-505. doi: 10.1136/amiajnl-2011-000217
- Bradley, C. A., Rolka, H., Walker, D., & Loonsk, J. (2005). BioSense: implementation of a National Early Event Detection and Situational Awareness System. *MMWR Morb Mortal Wkly Rep*, 54 Suppl, 11-19.
- Bronnert, J., Daube, J., Jopp, G., Peterson, K., Rihanek, T., Scichilone, R., & Tucker, V. (2014). Optimizing data representation through the use of SNOMED CT. *J AHIMA*, 85(3), 48-50.
- Brownstein, J. S., & Freifeld, C. C. (2007). HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill*, 12(11), E071129 071125.
- Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection-- harnessing the Web for public health surveillance. *N Engl J Med*, 360(21), 2153-2155, 2157. doi: 10.1056/NEJMp0900702
- Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*, 5(7), e151. doi: 10.1371/journal.pmed.0050151

- Buckeridge, D. L., Burkom, H., Moore, A., Pavlin, J., Cutchis, P., & Hogan, W. (2004). Evaluation of syndromic surveillance systems--design of an epidemic simulation model. *MMWR Morb Mortal Wkly Rep*, *53 Suppl*, 137-143.
- Buehler, J. W., Centers for Disease, Control, & Prevention. (2012). CDC's vision for public health surveillance in the 21st century. Introduction. *MMWR Surveill Summ*, *61 Suppl*, 1-2.
- Buehler, J. W., Hopkins, R. S., Overhage, J. M., Sosin, D. M., Tong, V., & Group, C. D. C. Working. (2004). Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep*, *53(RR-5)*, 1-11.
- Buehler, J. W., Whitney, E. A., Smith, D., Prietula, M. J., Stanton, S. H., & Isakov, A. P. (2009). Situational uses of syndromic surveillance. *Biosecur Bioterror*, *7(2)*, 165-177. doi: 10.1089/bsp.2009.0013
- Building rotavirus laboratory capacity to support the Global Rotavirus Surveillance Network. (2013). *Wkly Epidemiol Rec*, *88(21)*, 217-223.
- Burr, T., Graves, T., Klamann, R., Michalak, S., Picard, R., & Hengartner, N. (2006). Accounting for seasonal patterns in syndromic surveillance data for outbreak detection. *BMC Med Inform Decis Mak*, *6*, 40. doi: 10.1186/1472-6947-6-40
- Canas, L. C., Lohman, K., Pavlin, J. A., Endy, T., Singh, D. L., Pandey, P., . . . Gaydos, J. C. (2000). The Department of Defense laboratory-based global influenza surveillance system. *Mil Med*, *165(7 Suppl 2)*, 52-56.

- Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*, 49(10), 1557-1564. doi: 10.1086/630200
- Carnevale, R. J., Talbot, T. R., Schaffner, W., Bloch, K. C., Daniels, T. L., & Miller, R. A. (2011). Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks. *J Am Med Inform Assoc*, 18(4), 466-472. doi: 10.1136/amiajnl-2011-000216
- Centers for Disease, Control, & Prevention. (2002). Syndromic surveillance for bioterrorism following the attacks on the World Trade Center--New York City, 2001. *MMWR Morb Mortal Wkly Rep*, 51 Spec No, 13-15.
- Centers for Disease, Control, & Prevention. (2010). Update: influenza activity - United States, 2009-10 season. *MMWR Morb Mortal Wkly Rep*, 59(29), 901-908.
- Centers for Disease, Control, & Prevention. (2011). Assessment of ESSENCE performance for influenza-like illness surveillance after an influenza outbreak--U.S. Air Force Academy, Colorado, 2009. *MMWR Morb Mortal Wkly Rep*, 60(13), 406-409.
- Centers for Disease, Control, & Prevention. (2013). Comparison of provisional with final notifiable disease case counts - National Notifiable Diseases Surveillance System, 2009. *MMWR Morb Mortal Wkly Rep*, 62(36), 747-751.
- Chen, J. H., Schmit, K., Chang, H., Herlihy, E., Miller, J., & Smith, P. (2005). Use of Medicaid prescription data for syndromic surveillance--New York. *MMWR Morb Mortal Wkly Rep*, 54 Suppl, 31-34.

- Choi, B. C. (2012). The past, present, and future of public health surveillance. *Scientifica (Cairo)*, 2012, 875253. doi: 10.6064/2012/875253
- Cohen, S. M., Gray, K. M., Ocfemia, M. C., Johnson, A. S., & Hall, H. I. (2014). The status of the National HIV Surveillance System, United States, 2013. *Public Health Rep*, 129(4), 335-341.
- Collier, N., Doan, S., Kawazoe, A., Goodwin, R. M., Conway, M., Tateno, Y., . . . Taniguchi, K. (2008). BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics*, 24(24), 2940-2941. doi: 10.1093/bioinformatics/btn534
- Communicable diseases surveillance. Presentation of NNDSS data. (2001). *Commun Dis Intell*, 25(1), 18-30.
- Cowen, P., Garland, T., Hugh-Jones, M. E., Shimshony, A., Handysides, S., Kaye, D., . . . Woodall, J. (2006). Evaluation of ProMED-mail as an electronic early warning system for emerging animal diseases: 1996 to 2004. *J Am Vet Med Assoc*, 229(7), 1090-1099. doi: 10.2460/javma.229.7.1090
- Cubillas, J. J., Ramos, M. I., Feito, F. R., & Urena, T. (2014). An improvement in the appointment scheduling in primary health care centers using data mining. *J Med Syst*, 38(8), 89. doi: 10.1007/s10916-014-0089-y
- Cunningham, C. T., Cai, P., Topps, D., Svenson, L. W., Jette, N., & Quan, H. (2014). Mining rich health data from Canadian physician claims: features and face validity. *BMC Res Notes*, 7, 682. doi: 10.1186/1756-0500-7-682
- De Florentiis, D., Parodi, V., Orsi, A., Rossi, A., Altomonte, F., Canepa, P., . . . Ansaldi, F. (2011). Impact of influenza during the post-pandemic season: epidemiological

- picture from syndromic and virological surveillance. *J Prev Med Hyg*, 52(3), 134-136.
- Dombrowski, J. C., Buskin, S. E., Bennett, A., Thiede, H., & Golden, M. R. (2014). Use of Multiple Data Sources and Individual Case Investigation to Refine Surveillance-Based Estimates of the HIV Care Continuum. *J Acquir Immune Defic Syndr*, 67(3), 323-330. doi: 10.1097/QAI.0000000000000302
- Dong, X., Bahroos, N., Sadhu, E., Jackson, T., Chukhman, M., Johnson, R., . . . Hynes, D. (2013). Leverage hadoop framework for large scale clinical informatics applications. *AMIA Jt Summits Transl Sci Proc*, 2013, 53.
- Dorea, F. C., Lindberg, A., McEwen, B. J., Revie, C. W., & Sanchez, J. (2014). Syndromic surveillance using laboratory test requests: a practical guide informed by experience with two systems. *Prev Vet Med*, 116(3), 313-324. doi: 10.1016/j.prevetmed.2014.04.001
- Dorea, F. C., McEwen, B. J., McNab, W. B., Sanchez, J., & Revie, C. W. (2013). Syndromic surveillance using veterinary laboratory data: algorithm combination and customization of alerts. *PLoS One*, 8(12), e82183. doi: 10.1371/journal.pone.0082183
- Espino, J. U., Wagner, M. M., Tsui, F. C., Su, H. D., Olszewski, R. T., Lie, Z., . . . Dara, J. (2004). The RODS Open Source Project: removing a barrier to syndromic surveillance. *Stud Health Technol Inform*, 107(Pt 2), 1192-1196.
- Farran, B., Channanath, A. M., Behbehani, K., & Thanaraj, T. A. (2013). Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity:

- machine-learning algorithms and validation using national health data from Kuwait--a cohort study. *BMJ Open*, 3(5). doi: 10.1136/bmjopen-2012-002457
- Fearnhead, P., Giagos, V., & Sherlock, C. (2014). Inference for reaction networks using the linear noise approximation. *Biometrics*, 70(2), 457-466. doi: 10.1111/biom.12152
- Fidahussein, M., & Vreeman, D. J. (2014). A corpus-based approach for automated LOINC mapping. *J Am Med Inform Assoc*, 21(1), 64-72. doi: 10.1136/amiajnl-2012-001159
- Freifeld, C. C., Mandl, K. D., Reis, B. Y., & Brownstein, J. S. (2008). HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc*, 15(2), 150-157. doi: 10.1197/jamia.M2544
- Fricke, R. D., Jr., Hegler, B. L., & Dunfee, D. A. (2008). Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology. *Stat Med*, 27(17), 3407-3429. doi: 10.1002/sim.3197
- Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S., Park, D. J., Kanneh, L., . . . Sabeti, P. C. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, 345(6202), 1369-1372. doi: 10.1126/science.1259657
- Goedert, J. (2008). Wishing for a better EHR. Electronic health records users share their wish lists for added functionality. *Health Data Manag*, 16(9), 26-28, 30, 32 passim.

- Golder, D. (2010). Meaningful use of electronic health records (EHR)--the rules are final. Now what? *J Okla State Med Assoc*, 103(9), 433.
- Gotz, D., Wang, F., & Perer, A. (2014). A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *J Biomed Inform*, 48, 148-159. doi: 10.1016/j.jbi.2014.01.007
- Gould, L. H., Rosenblum, I., Nicholas, D., Phan, Q., & Jones, T. F. (2013). Contributing factors in restaurant-associated foodborne disease outbreaks, FoodNet sites, 2006 and 2007. *J Food Prot*, 76(11), 1824-1828. doi: 10.4315/0362-028X.JFP-13-037
- Guasticchi, G., Giorgi Rossi, P., Lori, G., Genio, S., Biagetti, F., Gabriele, S., . . . Borgia, P. (2009). Syndromic surveillance: sensitivity and positive predictive value of the case definitions. *Epidemiol Infect*, 137(5), 662-671. doi: 10.1017/S0950268808001374
- Hall, R. L., Jones, J. L., Hurd, S., Smith, G., Mahon, B. E., & Herwaldt, B. L. (2012). Population-based active surveillance for Cyclospora infection--United States, Foodborne Diseases Active Surveillance Network (FoodNet), 1997-2009. *Clin Infect Dis*, 54 Suppl 5, S411-417. doi: 10.1093/cid/cis049
- Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Med*, 10(4), e1001413. doi: 10.1371/journal.pmed.1001413
- Hayrinen, K., & Mykkanen, J. (2011). Mapping the Finnish National EHR to the LOINC. *Stud Health Technol Inform*, 165, 161-166.

- Henao, O. L., Crim, S. M., & Hoekstra, R. M. (2012). Calculating a measure of overall change in the incidence of selected laboratory-confirmed infections with pathogens transmitted commonly through food in the Foodborne Diseases Active Surveillance Network (FoodNet), 1996-2010. *Clin Infect Dis*, *54 Suppl 5*, S418-420. doi: 10.1093/cid/cis244
- Henry, J. V., Magruder, S., & Snyder, M. (2004). Comparison of office visit and nurse advice hotline data for syndromic surveillance--Baltimore-Washington, D.C., metropolitan area, 2002. *MMWR Morb Mortal Wkly Rep*, *53 Suppl*, 112-116.
- Heymann, D. L., & Rodier, G. R. (1998). Global surveillance of communicable diseases. *Emerg Infect Dis*, *4*(3), 362-365. doi: 10.3201/eid0403.980305
- Holtry, R. S., Hung, L. M., & Lewis, S. H. (2010). Utility of the ESSENCE Surveillance System in Monitoring the H1N1 Outbreak. *Online J Public Health Inform*, *2*(3). doi: 10.5210/ojphi.v2i3.3028
- Howard, J. (2000). Communicable disease and infection control: the surveillance contribution. *Br J Community Nurs*, *5*(12), 619-620, 622-617.
- Hu, Y. W., Kinsler, J. J., Sheng, Z., Kang, T., Bingham, T., & Frye, D. M. (2012). Using laboratory surveillance data to estimate engagement in care among persons living with HIV in Los Angeles County, 2009. *AIDS Patient Care STDS*, *26*(8), 471-478. doi: 10.1089/apc.2011.0371
- Hunscher, D., Boyd, A., Green, L. A., & Clauw, D. J. (2006). Representing natural-language case report form terminology using Health Level 7 Common Document Architecture, LOINC, and SNOMED-CT: lessons learned. *AMIA Annu Symp Proc*, 961.



- Hurd, S., Patrick, M., Hatch, J., Clogher, P., Wymore, K., Cronquist, A. B., . . . Fitzgerald, C. (2012). Clinical laboratory practices for the isolation and identification of *Campylobacter* in Foodborne Diseases Active Surveillance Network (FoodNet) sites: baseline information for understanding changes in surveillance data. *Clin Infect Dis*, *54 Suppl 5*, S440-445. doi: 10.1093/cid/cis245
- Jajosky, R. A., & Groseclose, S. L. (2004). Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health*, *4*, 29. doi: 10.1186/1471-2458-4-29
- Jeremy Sueker, J., Blazes, D. L., Johns, M. C., Blair, P. J., Sjoberg, P. A., Tjaden, J. A., . . . Do, D. Influenza Working Group. (2010). Influenza and respiratory disease surveillance: the US military's global laboratory-based network. *Influenza Other Respir Viruses*, *4*(3), 155-161. doi: 10.1111/j.1750-2659.2010.00129.x
- Jhung, M. A., D'Mello, T., Perez, A., Aragon, D., Bennett, N. M., Cooper, T., . . . Chaves, S. S. (2014). Hospital-onset influenza hospitalizations--United States, 2010-2011. *Am J Infect Control*, *42*(1), 7-11. doi: 10.1016/j.ajic.2013.06.018
- Johnson, L. R., Gould, L. H., Dunn, J. R., Berkelman, R., Mahon, B. E., & Foodnet Travel Working, Group. (2011). Salmonella infections associated with international travel: a Foodborne Diseases Active Surveillance Network (FoodNet) study. *Foodborne Pathog Dis*, *8*(9), 1031-1037. doi: 10.1089/fpd.2011.0854
- Johnson, M. G., Williams, J., Lee, A., & Bradley, K. K. (2014). Completeness and timeliness of electronic vs. conventional laboratory reporting for communicable disease surveillance--Oklahoma, 2011. *Public Health Rep*, *129*(3), 261-266.

- Johnson, N. B., Hayes, L. D., Brown, K., Hoo, E. C., & Ethier, K. A. (2014). CDC National Health Report: Leading Causes of Morbidity and Mortality and Associated Behavioral Risk and Protective Factors-United States, 2005-2013. *MMWR Surveill Summ*, 63, 3-27.
- Joseph, S., Snow, M., Furukawa, M., Posnack, S., & Chaffee, M. A. (2014). HITECH spurs EHR vendor competition and innovation, resulting in increased adoption. *Am J Manag Care*, 20(9), 734-740.
- Karami, M. (2012). Validity of evaluation approaches for outbreak detection methods in syndromic surveillance systems. *Iran J Public Health*, 41(11), 102-103.
- Kashiouris, M., O'Horo, J. C., Pickering, B. W., & Herasevich, V. (2013). Diagnostic performance of electronic syndromic surveillance systems in acute care: a systematic review. *Appl Clin Inform*, 4(2), 212-224. doi: 10.4338/ACI-2012-12-RA-0053
- Katz, R., May, L., Baker, J., & Test, E. (2011). Redefining syndromic surveillance. *J Epidemiol Glob Health*, 1(1), 21-31. doi: 10.1016/j.jegh.2011.06.003
- Kaufman, Z., & Shohat, T. (2014). [Syndromic surveillance in public health practice--an updated view]. *Harefuah*, 153(3-4), 188-191, 238, 237.
- Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., & Stahl, D. (2013). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Stat Methods Med Res*. doi: 10.1177/0962280213502437
- Lanham, H. J., Leykum, L. K., & McDaniel, R. R., Jr. (2012). Same organization, same electronic health records (EHRs) system, different use: exploring the linkage

- between practice member communication patterns and EHR use patterns in an ambulatory care setting. *J Am Med Inform Assoc*, 19(3), 382-391. doi: 10.1136/amiajnl-2011-000263
- Lazarus, R., Kleinman, K. P., Dashevsky, I., DeMaria, A., & Platt, R. (2001). Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *BMC Public Health*, 1, 9.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science*, 343(6176), 1203-1205. doi: 10.1126/science.1248506
- Lee, J., Hyeon, J. W., Kim, S. Y., Hwang, K. J., Ju, Y. R., & Ryou, C. (2014). Review: Laboratory diagnosis and surveillance of Creutzfeldt-Jakob disease. *J Med Virol*. doi: 10.1002/jmv.24004
- Lee, K. K., Tang, W. C., & Choi, K. S. (2013). Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage. *Comput Methods Programs Biomed*, 110(1), 99-109. doi: 10.1016/j.cmpb.2012.10.018
- Lee, L. H., Gross, A., Hartung, M., Liou, D. M., & Rahm, E. (2014). A multi-part matching strategy for mapping LOINC with laboratory terminologies. *J Am Med Inform Assoc*, 21(5), 792-800. doi: 10.1136/amiajnl-2013-002139
- Lee, L. M., Thacker, S. B., Centers for Disease, Control, & Prevention. (2011). The cornerstone of public health practice: public health surveillance, 1961--2011. *MMWR Surveill Summ*, 60 Suppl 4, 15-21.

- Lemay, R., Mawudeku, A., Shi, Y., Ruben, M., & Achonu, C. (2008). Syndromic surveillance for influenzalike illness. *Bio Secur Bioterror*, 6(2), 161-170. doi: 10.1089/bsp.2007.0056
- Lombardo, J., Burkom, H., Elbert, E., Magruder, S., Lewis, S. H., Loschen, W., . . . Pavlin, J. (2003). A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). *J Urban Health*, 80(2 Suppl 1), i32-42.
- Lombardo, J. S., Burkom, H., & Pavlin, J. (2004). ESSENCE II and the framework for evaluating syndromic surveillance systems. *MMWR Morb Mortal Wkly Rep*, 53 Suppl, 159-165.
- Loonsk, J. W. (2004). BioSense--a national initiative for early detection and quantification of public health emergencies. *MMWR Morb Mortal Wkly Rep*, 53 Suppl, 53-55.
- Lucas, P. (2004). Bayesian analysis, pattern analysis, and data mining in health care. *Curr Opin Crit Care*, 10(5), 399-403.
- Lyon, A., Nunn, M., Grossel, G., & Burgman, M. (2012). Comparison of web-based biosecurity intelligence systems: BioCaster, EpiSPIDER and HealthMap. *Transbound Emerg Dis*, 59(3), 223-232. doi: 10.1111/j.1865-1682.2011.01258.x
- Ma, H., Rolka, H., Mandl, K., Buckeridge, D., Fleischauer, A., & Pavlin, J. (2005). Implementation of laboratory order data in BioSense Early Event Detection and Situation Awareness System. *MMWR Morb Mortal Wkly Rep*, 54 Suppl, 27-30.
- MacIntosh, V. H., Tastad, K. J., & Eick-Cost, A. A. (2013). Mid-season influenza vaccine effectiveness 2011-2012: a Department of Defense Global, Laboratory-based,

- Influenza Surveillance System case-control study estimate. *Vaccine*, 31(13), 1651-1655. doi: 10.1016/j.vaccine.2013.01.022
- Madoff, L. C. (2004). ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis*, 39(2), 227-232. doi: 10.1086/422003
- Madoff, L. C., & Woodall, J. P. (2005). The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch Med Res*, 36(6), 724-730. doi: 10.1016/j.arcmed.2005.06.005
- Manikonda, K., Palmer, A., Wymore, K., McMillian, M., Nicholson, C., Hurd, S., . . . Patrick, M. (2012). Validating deaths reported in the Foodborne Diseases Active Surveillance Network (FoodNet): are all deaths being captured? *Clin Infect Dis*, 54 Suppl 5, S421-423. doi: 10.1093/cid/cis266
- Marella, W. M., Sparnon, E., & Finley, E. (2014). Screening Electronic Health Record-Related Patient Safety Reports Using Machine Learning. *J Patient Saf*. doi: 10.1097/PTS.0000000000000104
- Matheny, M. E., Peterson, J. F., Eden, S. K., Hung, A. M., Speroff, T., Abdel-Kader, K., . . . Siew, E. D. (2014). Laboratory test surveillance following acute kidney injury. *PLoS One*, 9(8), e103746. doi: 10.1371/journal.pone.0103746
- Mathias, J. S., Agrawal, A., Feinglass, J., Cooper, A. J., Baker, D. W., & Choudhary, A. (2013). Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. *J Am Med Inform Assoc*, 20(e1), e118-124. doi: 10.1136/amiajnl-2012-001360
- McElwain, T. F. (2010). Laboratory preparedness: building a cornerstone for global surveillance. *Future Microbiol*, 5(4), 531-533. doi: 10.2217/fmb.10.26

- Mehta, K. A. (2010). Electronic Health Records (EHR) is coming. *J Okla State Med Assoc*, 103(9), 407.
- Milinovich, G. J., Williams, G. M., Clements, A. C., & Hu, W. (2014). Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis*, 14(2), 160-168. doi: 10.1016/S1473-3099(13)70244-5
- Miller, M. J., Degenholtz, H. B., Gazmararian, J. A., Lin, C. J., Ricci, E. M., & Sereika, S. M. (2007). Identifying elderly at greatest risk of inadequate health literacy: a predictive model for population-health decision makers. *Res Social Adm Pharm*, 3(1), 70-85. doi: 10.1016/j.sapharm.2006.06.001
- Mok, J., Ho, P. C., Tsui, K., Ng, L., & Fung, V. (2013). Quality assurance of LOINC mapping for laboratory tests - a local experience with people, process and technology. *Stud Health Technol Inform*, 192, 975.
- Mold, F., Ellis, B., de Lusignan, S., Sheikh, A., Wyatt, J. C., Cavill, M., . . . Rafi, I. (2012). The provision and impact of online patient access to their electronic health records (EHR) and transactional services on the quality and safety of health care: systematic review protocol. *Inform Prim Care*, 20(4), 271-282.
- Morse, S. S. (2012). Public health surveillance and infectious disease detection. *Biosecur Bioterror*, 10(1), 6-16. doi: 10.1089/bsp.2011.0088
- Mu, A., Rutledge, J., Mills, P., & Paul, S. (2014). Incidence of Kaposi Sarcoma and Associated Mortality in Fresno, California, 1998 to 2012. *J Int Assoc Provid AIDS Care*. doi: 10.1177/2325957414531454
- National Electronic Disease Surveillance System Working, Group. (2001). National Electronic Disease Surveillance System (NEDSS): a standards-based approach

- to connect public health and clinical medicine. *J Public Health Manag Pract*, 7(6), 43-50.
- Naumova, E. N., O'Neil, E., & MacNeill, I. (2005). INFERNO: a system for early outbreak detection and signature forecasting. *MMWR Morb Mortal Wkly Rep*, 54 Suppl, 77-83.
- Niesters, H. G., Rossen, J. W., van der Avoort, H., Baas, D., Benschop, K., Claas, E. C., . . . Koopmans, M. (2013). Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories. *Euro Surveill*, 18(4), 20387.
- Ningthoujam, S. S., Choudhury, M. D., Potsangbam, K. S., Chetia, P., Nahar, L., Sarker, S. D., . . . Talukdar, A. D. (2014). NoSQL Data Model for Semi-automatic Integration of Ethnomedicinal Plant Data from Multiple Sources. *Phytochem Anal*, 25(6), 495-507. doi: 10.1002/pca.2520
- Nordin, J. D., Goodman, M. J., Kulldorff, M., Ritzwoller, D. P., Abrams, A. M., Kleinman, K., . . . Platt, R. (2005). Simulated anthrax attacks and syndromic surveillance. *Emerg Infect Dis*, 11(9), 1394-1398. doi: 10.3201/eid1109.050223
- Nsubuga, P., White, M. E., Thacker, S. B., Anderson, M. A., Blount, S. B., Broome, C. V., . . . Trostle, M. (2006). Public Health Surveillance: A Tool for Targeting and Monitoring Interventions. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills & P. Musgrove (Eds.), *Disease Control Priorities in Developing Countries* (2nd ed.). Washington (DC).
- O'Brien, S. J., & Christie, P. (1997). Do CuSums have a role in routine communicable disease surveillance? *Public Health*, 111(4), 255-258.

- O'Connell, E. K., Zhang, G., Leguen, F., Llau, A., & Rico, E. (2010). Innovative uses for syndromic surveillance. *Emerg Infect Dis*, *16*(4), 669-671. doi: 10.3201/eid1604.090688
- O'Connor, M. J., Buckeridge, D. L., Choy, M., Crubezy, M., Pincus, Z., & Musen, M. A. (2003). BioSTORM: a system for automated surveillance of diverse data sources. *AMIA Annu Symp Proc*, 1071.
- Paraiso, S., Perez Del Rey, D., Bucur, A., Claerhout, B., & Alonso-Calvo, R. (2014). Semantic normalization and query abstraction based on SNOMED-CT and HL7: Supporting multi-centric clinical trials. *IEEE J Biomed Health Inform*. doi: 10.1109/JBHI.2014.2357025
- Partington, S. N., Papakroni, V., & Menzies, T. (2014). Optimizing data collection for public health decisions: a data mining approach. *BMC Public Health*, *14*, 593. doi: 10.1186/1471-2458-14-593
- Patterson-Lomba, O., Van Noort, S., Cowling, B. J., Wallinga, J., Gomes, M. G., Lipsitch, M., & Goldstein, E. (2014). Utilizing syndromic surveillance data for estimating levels of influenza circulation. *Am J Epidemiol*, *179*(11), 1394-1401. doi: 10.1093/aje/kwu061
- Pattie, D. C., Atherton, M. J., & Cox, K. L. (2009). Electronic influenza monitoring: evaluation of body temperature to classify influenza-like illness in a syndromic surveillance system. *Qual Manag Health Care*, *18*(2), 91-102. doi: 10.1097/QMH.0b013e3181a0274d



- Patwardhan, A., & Bilkovski, R. (2012). Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. *PLoS One*, 7(8), e43611. doi: 10.1371/journal.pone.0043611
- Paz-Bailey, G., Raymond, H. F., Lansky, A., & Mermin, J. (2014). Using the National HIV Behavioral Surveillance System to inform HIV prevention efforts in the United States. *AIDS Behav*, 18 Suppl 3, S233-236. doi: 10.1007/s10461-014-0738-7
- Perry, A. G., Korenberg, M. J., Hall, G. G., & Moore, K. M. (2011). Modeling and syndromic surveillance for estimating weather-induced heat-related illness. *J Environ Public Health*, 2011, 750236. doi: 10.1155/2011/750236
- Pollack, M. P., Pringle, C., Madoff, L. C., & Memish, Z. A. (2013). Latest outbreak news from ProMED-mail: novel coronavirus -- Middle East. *Int J Infect Dis*, 17(2), e143-144. doi: 10.1016/j.ijid.2012.12.001
- Rabon-Stith, K. M., McGuinness, C. B., Saunders, B., Edelman, L., Kumar, V. R., & Boron, M. L. (2013). Laboratory testing trends for respiratory syncytial virus, 2007-2011. *J Clin Virol*, 58(3), 575-578. doi: 10.1016/j.jcv.2013.09.012
- Ranallo, P. A., Adam, T. J., Nelson, K. J., Krueger, R. F., LaVenture, M., & Chute, C. G. (2013). Psychological assessment instruments: a coverage analysis using SNOMED CT, LOINC and QS terminology. *AMIA Annu Symp Proc*, 2013, 1333-1340.
- Reis, B. Y., & Mandl, K. D. (2004). Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med*, 44(3), 235-241. doi: 10.1016/S0196064404003117

- Respass, R. A., Rayfield, M. A., & Dondero, T. J. (2001). Laboratory testing and rapid HIV assays: applications for HIV surveillance in hard-to-reach populations. *AIDS, 15 Suppl 3*, S49-59.
- Richardson, P., Henderson, L., Davila, J. A., Kramer, J. R., Fitton, C. P., Chen, G. J., & El-Serag, H. B. (2010). Surveillance for hepatocellular carcinoma: development and validation of an algorithm to classify tests in administrative and laboratory data. *Dig Dis Sci, 55*(11), 3241-3251. doi: 10.1007/s10620-010-1387-y
- Robinson, S. (2014). Lyme Disease in Maine: a Comparison of NEDSS Surveillance Data and Maine Health Data Organization Hospital Discharge data. *Online J Public Health Inform, 5*(3), 231. doi: 10.5210/ojphi.v5i3.4990
- Sabharwal, C. J., Braunstein, S. L., Robbins, R. S., & Shepard, C. W. (2014). Optimizing the use of surveillance data for monitoring the care status of persons recently diagnosed with HIV in NYC. *J Acquir Immune Defic Syndr, 65*(5), 571-578. doi: 10.1097/QAI.0000000000000077
- Salvadores, M., Alexander, P. R., Musen, M. A., & Noy, N. F. (2013). BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semant Web, 4*(3), 277-284.
- Samoff, E., Fangman, M. T., Hakenewerth, A., Ising, A., & Waller, A. E. (2014). Use of syndromic surveillance at local health departments: movement toward more effective systems. *J Public Health Manag Pract, 20*(4), E25-30. doi: 10.1097/PHH.0b013e3182a505ac

- Santillana, M., Zhang, D. W., Althouse, B. M., & Ayers, J. W. (2014). What can digital disease detection learn from (an external revision to) google flu trends? *Am J Prev Med*, *47*(3), 341-347. doi: 10.1016/j.amepre.2014.05.020
- Savory, D. J., Cox, K. L., Emch, M., Alemi, F., & Pattie, D. C. (2010). Enhancing spatial detection accuracy for syndromic surveillance with street level incidence data. *Int J Health Geogr*, *9*, 1. doi: 10.1186/1476-072X-9-1
- Scallan, E., & Mahon, B. E. (2012). Foodborne Diseases Active Surveillance Network (FoodNet) in 2012: a foundation for food safety in the United States. *Clin Infect Dis*, *54 Suppl 5*, S381-384. doi: 10.1093/cid/cis257
- Scarpino, S. V., Dimitrov, N. B., & Meyers, L. A. (2012). Optimizing provider recruitment for influenza surveillance networks. *PLoS Comput Biol*, *8*(4), e1002472. doi: 10.1371/journal.pcbi.1002472
- Schirmer, P., Lucero, C., Oda, G., Lopez, J., & Holodniy, M. (2010). Effective detection of the 2009 H1N1 influenza pandemic in U.S. Veterans Affairs medical centers using a national electronic biosurveillance system. *PLoS One*, *5*(3), e9533. doi: 10.1371/journal.pone.0009533
- Shult, P. A., & Kirk, C. (2003). Laboratory-based surveillance for influenza: role of the Wisconsin State Laboratory of Hygiene. *WMJ*, *102*(6), 26-30.
- Sintchenko, V., & Gallego, B. (2009). Laboratory-guided detection of disease outbreaks: three generations of surveillance systems. *Arch Pathol Lab Med*, *133*(6), 916-925. doi: 10.1043/1543-2165-133.6.916

- Sittig, D. F., Singh, H., & Longhurst, C. A. (2013). Rights and responsibilities of electronic health records (EHR) users caring for children. *Arch Argent Pediatr*, 111(6), 468-471. doi: 10.1590/S0325-00752013000600003
- Smith, P. F., Hadler, J. L., Stanbury, M., Rolfs, R. T., Hopkins, R. S., & Group, Cste Surveillance Strategy. (2013). "Blueprint version 2.0": updating public health surveillance for the 21st century. *J Public Health Manag Pract*, 19(3), 231-239. doi: 10.1097/PHH.0b013e318262906e
- Sokolow, L. Z., Grady, N., Rolka, H., Walker, D., McMurray, P., English-Bullard, R., & Loonsk, J. (2005). Deciphering data anomalies in BioSense. *MMWR Morb Mortal Wkly Rep*, 54 Suppl, 133-139.
- Sosin, D. M. (2003a). Draft framework for evaluating syndromic surveillance systems. *J Urban Health*, 80(2 Suppl 1), i8-13.
- Sosin, D. M. (2003b). Syndromic surveillance: the case for skillful investment. *Biosecur Bioterror*, 1(4), 247-253. doi: 10.1089/153871303771861441
- Spyropoulos, A. C., Viscusi, A., Singhal, N., Gilleylen, J., Kouides, P., Howard, M., . . . Triller, D. M. (2014). Features of Electronic Health Records Necessary for the Delivery of Optimized Anticoagulant Therapy: Consensus of the EHR Task Force of the New York State Anticoagulation Coalition. *Ann Pharmacother*. doi: 10.1177/1060028014555176
- Stewart, A., & Denecke, K. (2010). Using ProMED-Mail and MedWorm blogs for cross-domain pattern analysis in epidemic intelligence. *Stud Health Technol Inform*, 160(Pt 1), 437-441.

- Sugawara, T., Ohkusa, Y., Shigematsu, M., Taniguchi, K., Murata, A., & Okabe, N. (2007). [An experimental study for syndromic surveillance using OTC sales]. *Kansenshogaku Zasshi*, 81(3), 235-241.
- Tabak, Y. P., Sun, X., Nunez, C. M., & Johannes, R. S. (2014). Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS). *J Am Med Inform Assoc*, 21(3), 455-463. doi: 10.1136/amiajnl-2013-001790
- Terry, K. (2010). Electronic health records. States setting high standards for EHR use. *Hosp Health Netw*, 84(7), 16.
- Thacker, S. B., Berkelman, R. L., & Stroup, D. F. (1989). The science of public health surveillance. *J Public Health Policy*, 10(2), 187-203.
- Thacker, S. B., Qualters, J. R., Lee, L. M., Centers for Disease, Control, & Prevention. (2012). Public health surveillance in the United States: evolution and challenges. *MMWR Surveill Summ*, 61 Suppl, 3-9.
- Ting, J. S., Tsang, A. H., Ip, A. W., & Ho, G. T. (2011). RF-Medisys: a radio frequency identification-based electronic medical record system for improving medical information accessibility and services at point of care. *HIM J*, 40(1), 25-32.
- Tokars, J. I., English, R., McMurray, P., & Rhodes, B. (2010). Summary of data reported to CDC's national automated biosurveillance system, 2008. *BMC Med Inform Decis Mak*, 10, 30. doi: 10.1186/1472-6947-10-30
- Uscher-Pines, L., Farrell, C. L., Cattani, J., Hsieh, Y. H., Moskal, M. D., Babin, S. M., . . . Rothman, R. E. (2009). A survey of usage protocols of syndromic surveillance

- systems by state public health departments in the United States. *J Public Health Manag Pract*, 15(5), 432-438. doi: 10.1097/PHH.0b013e3181a5d36b
- Valencia-Mendoza, A., & Bertozzi, S. M. (2008). A predictive model for the utilization of curative ambulatory health services in Mexico. *Salud Publica Mex*, 50(5), 397-407.
- van den Wijngaard, C. C., van Pelt, W., Nagelkerke, N. J., Kretzschmar, M., & Koopmans, M. P. (2011). Evaluation of syndromic surveillance in the Netherlands: its added value and recommendations for implementation. *Euro Surveill*, 16(9).
- Velardi, P., Stilo, G., Tozzi, A. E., & Gesualdo, F. (2014). Twitter mining for fine-grained syndromic surveillance. *Artif Intell Med*, 61(3), 153-163. doi: 10.1016/j.artmed.2014.01.002
- Vogt, R. L. (1996). Laboratory reporting and disease surveillance. *J Public Health Manag Pract*, 2(4), 28-30.
- Wang, T., Wang, Y., & Biedermann, S. (2013). Funding alternatives in EHR adoption: beyond HITECH incentives and traditional approaches. *Healthc Financ Manage*, 67(5), 86-91.
- Wang, Z., Shah, A. D., Tate, A. R., Denaxas, S., Shawe-Taylor, J., & Hemingway, H. (2012). Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PLoS One*, 7(1), e30412. doi: 10.1371/journal.pone.0030412

Wojcik, O. P., Brownstein, J. S., Chunara, R., & Johansson, M. A. (2014). Public health for the people: participatory infectious disease surveillance in the digital age.

*Emerg Themes Epidemiol*, 11, 7. doi: 10.1186/1742-7622-11-7

Worden, K., & Manson, G. (2007). The application of machine learning to structural health monitoring. *Philos Trans A Math Phys Eng Sci*, 365(1851), 515-537. doi: 10.1098/rsta.2006.1938

Xing, J., Burkom, H., & Tokars, J. (2011). Method selection and adaptation for distributed monitoring of infectious diseases for syndromic surveillance. *J Biomed Inform*, 44(6), 1093-1101. doi: 10.1016/j.jbi.2011.08.012

Yang, S. (1998). FoodNet and Enter-net: emerging surveillance programs for foodborne diseases. *Emerg Infect Dis*, 4(3), 457-458. doi: 10.3201/eid0403.980330

Youmans, E., Tripathi, A., Gibson, J. J., Stephens, T., & Duffus, W. A. (2011). Demographic characteristics and behavioral risk factors of HIV infection and association with survival among individuals 50 years or older. *South Med J*, 104(10), 669-675. doi: 10.1097/SMJ.0b013e31822dd3d8

Yuan, Y., Van Allen, E. M., Omberg, L., Wagle, N., Amin-Mansour, A., Sokolov, A., . . .

Liang, H. (2014). Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat Biotechnol*, 32(7), 644-652. doi: 10.1038/nbt.2940

Zikos, D., & Diomidous, M. (2012). Integration of data analysis methods in syndromic surveillance systems. *Stud Health Technol Inform*, 180, 1114-1116.