Synthesis of Research on the Value of Unconventional Data for Early Detection of Disease Outbreaks

Final Report

DARPA BioALIRT Program
October 2004

Compiled by:

Michael Wagner MD, PhD
Associate Professor of Medicine and Intelligent Systems
Director, Real-time Outbreak and Disease Surveillance Laboratory
100 Technology Drive
University of Pittsburgh
Pittsburgh, PA 15219

Julie Pavlin, MD MPH
Chief, Department of Field Studies
Walter Reed Army Institute of Research
503 Robert Grant Ave
Silver Spring, MD 20910

Judith C. Brillman MD
Professor
Department of Emergency Medicine
University of New Mexico School of Medicine
Albuquerque NM

Douglas Stetson MD
General Dynamics Advanced Information Systems, Inc.
1400 Key Blvd, Ste 100
Arlington, VA 22209

Steve Magruder, Ph. D.
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Murray Campbell, Ph.D.
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Contributors

Steven M. Babin, MD, PhD
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Ronald M. Bloom
General Dynamics Advanced Information Systems
5933 West Century Blvd Suite 700
Los Angeles, CA 90045

James Brinkman
J&P Enterprises
616 Solar Road NW
Albuquerque, NM 87107

Wendy Chapman, PhD
RODS Laboratory
University of Pittsburgh

Karen Cheng, PhD
General Dynamics Advanced Information Systems, Inc.

Jeremy Espino, MD
RODS Laboratory
University of Pittsburgh

Elaine Florio, BS
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Virginia Foster, PhD MPH
Walter Reed Army Institute of Research
Senior Epidemiologist, Dpt of Field Studies
Walter Reed Army Institute of Research
503 Robert Grant Ave.
Silver Spring, MD 20910

William Hogan, MD, MS
Assistant Professor of Medicine
RODS Laboratory
University of Pittsburgh

Heather Johnson
RODS Laboratory
University of Pittsburgh

Andrew Kress
President
Surveillance Data, Inc.
220 W. Germantown Pike
Suite 140
Plymouth Meeting, PA 19462

Joe Lombardo, MS
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Gaurong Shah, MD, MPH
RODS Laboratory
University of Pittsburgh

David Siegrist, MA, MSM
Senior Research Fellow
Potomac Institute for Policy Studies
Arlington, VA 22203

Garrick Wallstrom, PhD
Instructor in Medicine
RODS Laboratory
University of Pittsburgh

Contractors

The ESSENCE Team (Johns Hopkins University Applied Physics Laboratory and the Walter Reed Army Institute of Research)

General Dynamics-Stanford University

IBM T.J. Watson Research Center

Mellon-Pitt Team (Carnegie Mellon University and the University of Pittsburgh)

Executive Summary

The DARPA Biological Advanced Leading Indicator Recognition Technology (Bio-ALIRT) program was a three-year program with a goal of improving the time of detection of bioterrorism events.

The four research teams in the Bio-ALIRT program were General Dynamics–Stanford University, IBM Corporation, Johns Hopkins University Applied Physics Laboratory–Walter Reed Army Institute of Research, and University of Pittsburgh–Carnegie Mellon University.

The two objectives of the program were to explore of the utility of novel types of surveillance data and develop new algorithms for the analysis of surveillance data.

This report summarizes the research on the utility of novel types of surveillance data.

The four teams examined data from 100 sources representing at least 20 different types of data

The positive findings were:

- Over-the-Counter (OTC) pharmaceutical sales were found to be a promising potential source of information about community illness. In particular, evidence was found that increases in pediatric electrolyte sales precede seasonal outbreaks of respiratory and diarrheal illness in children under five, that sales of pediatric cough and cold products and antipyretics correlate with influenza outbreaks, that flu medications precede seasonal outbreaks of respiratory illness, and that chest rub sales are highly correlated with the seasonal cycle of respiratory illness.

- Significant correlations were also found between non-seasonal fluctuations in OTC sales and clinical measures of community illness, but these results were not conclusive concerning a possible time-lead of the OTC fluctuations over clinical manifestations of illness.

- Advice-line call data were also found to be a useful and timely predictor of community illness. Nurse-advice / appointment calls categorized as acute respiratory illness were found to correlate with physician diagnosis of respiratory illness in the majority of cases, with the call preceding the physician encounters by a median of 5 hours. Telephone triage calls categorized as respiratory illness were found to precede influenza surveillance data by weeks.

- Automatically parsed emergency department chief complaints were found to correctly predict patient chart review diagnosis categories with modest reliability. When aggregated into community counts, the parsed chief complaint signal accurately detected winter outbreaks of respiratory and diarrheal illness in adults and children.

- Validations of military ICD9-coded physician diagnoses, reported soon after patient encounters have found that this promptly available data accurately predict outcomes of

more thorough reviews of patient charts (average sensitivity of 75%, 93% and 87% for respiratory, gastrointestinal and fever respectively).

- Automatic processing of free-text chest radiograph descriptions accurately predicted physician interpretation of the same reports for the findings of mediastinal widening and pneumonia.

- Preliminary evidence was found to indicate that self-assessed health surveys and telephone calls to medical offices anticipate physician office visits by three and four days respectively. These types of data sources may be available in selected environments, such as worksites, university campuses, or military bases.

- Fluctuations in elementary school absenteeism were found to correlate (correlations of 50% to 80%) with emergency room discharge diagnoses for acute illness.

- A sudden increase in visits to a hospital and clinic were found to be detectable by monitoring use of the parking facility.

Negative findings:

- No lead time was identified for influenza test requests, as compared to physician visit ICD9-coded diagnoses of influenza-like illness.

- Poison center data showed no disease effect during influenza outbreaks.

- Correlations between missed orthodontist appointments, and hospital discharge diagnoses for lower respiratory tract infections were found to be low (27%)

The methodological contributions of the research included:

A filtered correlation technique to separate measurements of data source timeliness based on seasonal disease activity from measurements based specifically on non-seasonal activity.

An empirical clustering method to automatically derive useful aggregations of OTC products for syndromic surveillance.

Validated techniques for automatically parsing Emergency Department chief complaints and chest radiograph descriptions and placing them in appropriate syndromic categories.

Methods to reduce the effects of confounders in OTC data through the use of auxiliary data and modeling.

The publications and technical reports resulting from this research number 22 with additional reports in preparation. They are listed in the bibliography and attached as an Appendix.

Table of Contents

1 INTRODUCTION

This report summarizes results from the Biological Advanced Leading Indicator Recognition Technology program (Bio-ALIRT), which was funded by the Defense Advanced Research Projects Agency from 2001-2003. The overall goal of the program was to develop technology for early detection of a covert biological event through statistical analysis of non traditional data sources. This report summarizes the work performed to identify and evaluate <u>non traditional data sources</u>. In particular, the report discusses those data sources that were obtained and analyzed by the program and the results both positive and negative of the research. Separate reports summarize the work on new outbreak detection algorithms.

The four research teams in the Bio-ALIRT program were General Dynamics–Stanford University, IBM Corporation, Johns Hopkins University Applied Physics Laboratory–Walter Reed Army Institute of Research, and University of Pittsburgh–Carnegie Mellon University.

The premise motivating this research is that epidemics make early footprints in data that are being collected routinely for other purposes. It is known that the sick exhibit behaviors such as purchasing over-the-counter medications in dealing with symptoms of a new illness before seeking professional medical care. These behaviors may result in data transactions at cash registers and other locations creating data that are a potential new source of data for public health surveillance. Examples of these potentially early disease indicators include: sales of over-the-counter (OTC) health care products; calls to health advice lines; health surveys administered at work or school; absence from work or school; changes in use patterns of utilities like potable water, wastewater service, and electrical service; indicators of disease in animals; queries from health websites; use of public transit; sales in cafeterias; and even the utilization levels of parking facilities that service medical facilities. The interest in these non traditional sources derives from their potential to provide an extra margin of earliness of detection relative to clinical data, (their 'timeliness' advantage).

1.1 Scope

The scope of BioALIRT research on data was limited to non traditional data sources. Our working definition of a non traditional data source was a type of data that was not currently in use in public health surveillance (at the time the Bio-ALIRT research began).

The four research teams obtained non traditional data from a variety of sources such as hospitals, outpatient practices, transit systems, and retail pharmacies (see Table 1). A dataset might include daily counts of emergency department visits, ICD-9 coded diagnoses from outpatient visits, orders for laboratory tests, radiology reports, prescriptions, EMS calls, or sales of cough preparations. The typical dataset represented a multi-year period and a large geographical area.

The choice of which data to study was guided by review of prior research and a prior analysis of data needed for bioterrorism detection [2]. Table 1 lists the types of data that the contractors obtained. Table 1 also lists the kinds of conventional disease surveillance data that the contractors obtained to use as gold-standard measures of actual disease activity in the research, as well as a summary of the methods and findings.

1.2 Methods

The key research hypothesis driving this research is that data routinely collected by businesses such as hospitals, transit authorities, and pharmacies can provide an early indication of a bioterrorism event.

A significant methodological problem in testing this hypothesis is that bioterrorism events are nearly non existent. Therefore, the actual hypotheses being tested in this research are about detectability of diseases like Influenza that occur frequently and can serve as surrogates for bioterrorism diseases. Specifically, the contractors studied outbreaks of gastrointestinal or respiratory diseases that happened to occur in the locations and timeframes represented in the research datasets. These outbreaks are "experiments of nature" that were used to understand the effect of such epidemics on the non-traditional source. In many cases the outbreaks studied were not caused by a single known organism, but were annual winter outbreaks of diarrhea and respiratory diseases in pediatric and adult populations.

Because of the use of surrogate diseases in this research, extrapolation of the results to bioterrorism diseases is required. The topic of surrogates for bioterrorism diseases is discussed more fully in [3]

The evidence developed by the BioALIRT program came from both laboratory experiments on historical datasets as well as field experience with deployed systems.

1.2.1 Laboratory experiments

The laboratory experiments measured correlation, sensitivity, false alarm rate, and timeliness.

The experiments that measured timeliness also measured sensitivity and false alarm rate. In general, there is a trade-off between *timeliness* of a data source--defined as how early the data show a signal of an outbreak relative to the start of the outbreak--and the number of false alarms due to the lack of diagnostic specificity of the data. Clinical data, for example, provide specific information about the causative organism but this information is available too late to mitigate many bioterrorism threats [4]. In contrast, preclinical data may provide a needed earlier signal, but at a cost of decreased specificity and resulting false alarms.

When evaluating new types of data, it is important to understand this trade off between timeliness and diagnostic specificity. Figure 1.1 characterizes many of the data sources studied along these two dimensions. Most early nonspecific indicators cannot be relied upon individually for outbreak detection but in concert with multiple data streams and two-stage approaches to detection, may increase the timeliness of outbreak detection.

The specific experimental methods used in this research are described in detail in subsequent sections. Briefly, researchers constructed time series of non traditional data and time series of traditional data and used correlation analysis to measure the time lags between the two series. They also used what we term the *Detection Algorithm Method* to measure the sensitivity, false alarm rate, and timeliness achievable from non traditional data relative to detection from more traditional public health surveillance data. Additionally researchers measured the diagnostic sensitivity and specificity of data directly by comparison with reference diagnoses for individual cases.

Table 1.1 Summary of Types of Data, Diseases, Experimental Methods, Results, and References

Data Type	Gold Standard	Syndrome or Diagnosis	Algorithm	Correlation	Sensitivity	Specificity	Timeliness (95% CI)	Ref
Pediatric electrolyte Sales	discharge diagnosis	Pediatric Respiratory and diarrhea	EMWA		1.0	1.0	1.7 weeks (0.5,2.9)	Hogan 2003
Flu meds and chest rubs sales	Outpatient ICD-9 codes	Respiratory	Correlation	0.895			2.9 days	Magruder 2003
Telephone Triage advice lines	CDC influenza data	Respiratory	CCF				1-5 weeks	Espino 2003
HMO appointment requests	Discharge diagnoses						5 hours	Henry
Automatically processed chief complaints	P&I discharge diagnoses	Adult Respiratory	ODAM		100	100	11 days (-10, 33)	Ivanov 2003
Automatically processed chief complaints	Pediatric P&I, RSV discharge diagnoses	Pediatric respiratory	ODAM		100	100	10 (-1.5,35)	Ivanov 2003
ICD-9 codes	Manual review of ED assessment	Respiratory GI Fever			0.76-0.9	0.94		Betancourt
Chest Radiographs	Manual review of reports	Anthrax	Key word based detector		0.86	0.99		Chapman 2003
Health Surveys	Outpatient ICD-9 codes	Respiratory	Correlation		0.82		3 days prior to office visit	Li
Telephone calls to medical providers	Outpatient ICD-9 codes	Respiratory	Correlation		0.72		4 days prior to office visit	
Influenza test orders	Outpatient ILI ICD-9 diagnosis	Influenza	Correlation	high			None	Chotani, Lewis, Coberly
Elementary school absenteeism	Outpatient and ED ICD-9 discharge diagnoses	All acute illnesses	Correlation	0.5-.08			2 days	Vasholz
Prescription drugs, military	Outpatient ICD-9 codes	Respiratory and GI	Correlation	high			None	Foster Elbert
Accession of influenza articles via the web	ILI reports by sentinel physicians reported to CDC	Influenza	Correlation CCF	0.67-0.8			-3 to 2 weeks	Johnson 2004
Transportation	Weekly regional percentage of positive influenza results	Respiratory	Correlation CCF	Poor, Moderate for seniors			4 weeks	Shah
Cafeteria beverage sales	Outpatient ICD-9 codes	Respiratory	correlation	0.66			19 days	IBM
Counts of users of medical facility parking structure	Baseline parking count during known outbreak		Exceed 1.65 SD		100%	100%		GD Cheng
Missed orthodontic appointments	Pediatric P&I, RSV discharge diagnoses	Respiratory	CCF				-3 weeks	Shah 2003

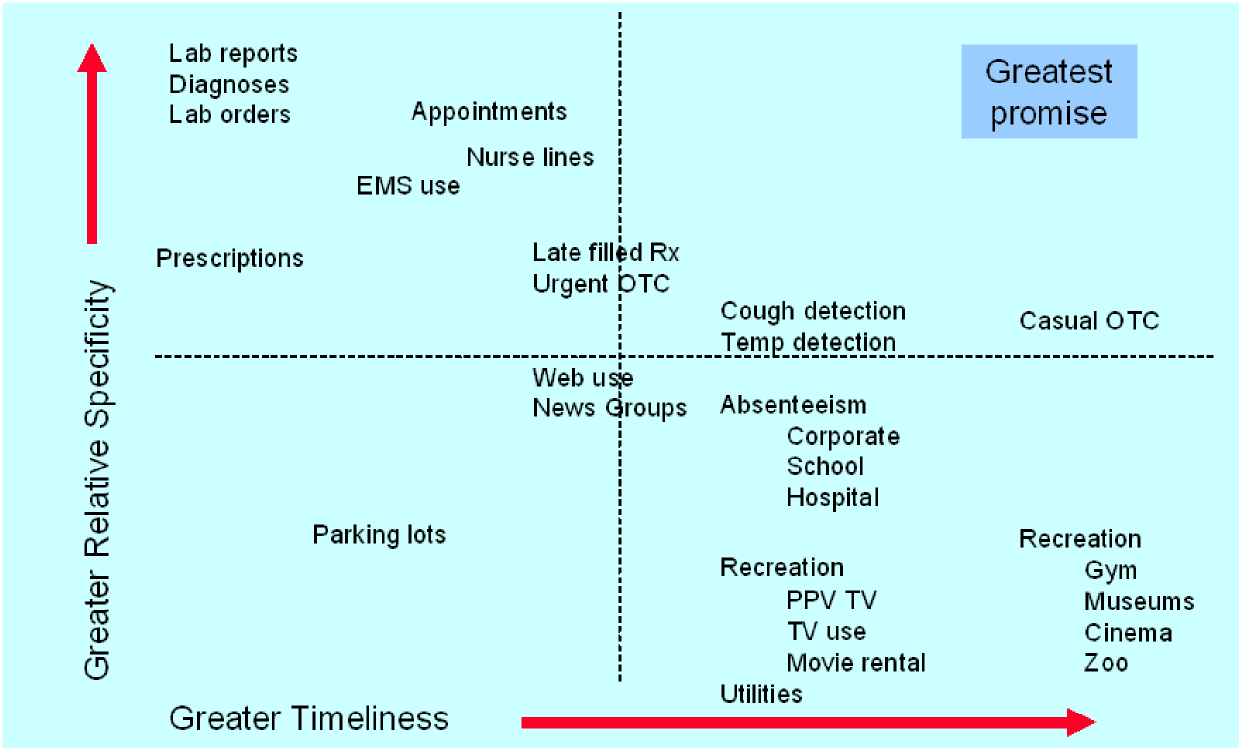

Figure 1.1 Testable hypotheses about the relative specificity and timeliness of non traditional data sources.

1.2.2 Field Testing

During the course of the BioALIRT program, the ESSENCE and RODS systems were in production operation in the National Capital Area, Pennsylvania, and Utah, and the National Retail Data Monitor came into existence in the last year of the program monitoring the entire United States. These systems both detected and failed to detect known outbreaks and these detections (and failures) serve as the highest level of confirmation (and refutation) of the underlying hypotheses for various types of data and outbreaks.

1.3 Overview of the Sections

In Section 2 we discuss gold standard datasets that were obtained and used in this research. These datasets include ICD-9-coded hospital discharge diagnoses, ICD-9-coded outpatient billing diagnoses, influenza surveillance data and hospital charts.

In Section 3 we discuss the types of data that research suggests are "leading indicators"--those considered most promising by the authors of this report. Data were considered promising based on:

- The weight of evidence accumulated by the Bio-ALIRT program
- The timeliness and availability of the data
- The weight of supporting evidence in the literature

For each data source, we discuss data characteristics such as spatial and temporal granularity, availability and latency. If relevant or available, we describe how some types of data must be grouped into analytic categories prior to further processing by detection algorithms. We also discuss results about known confounders and covariates that can be used to remove noise from the data. Most importantly, we synthesize the available evidence about detectability (size of outbreak detectable) and earliness to form conclusions about the state-of-the art for each indicator.

Section 4 discusses results for other data sources that were studied by Bio-ALIRT as potential timely indicators of disease. We used the same criteria for evaluation for these sources as in Section 3.

Section 5 discusses types of data that were identified by BioALIRT as having potential, and that should be considered in future research.

The published papers and technical reports from BioALIRT are included in Appendix.

2 GOLD STANDARD SURVEILLANCE DATA

This section describes datasets used by the research teams to identify the occurrence and timing of actual outbreaks of disease. We use the term "gold standard surveillance data" because we used these data to establish the existence and timing of outbreaks; that is, to establish the ground truth against which the detectability characteristics (sensitivity, false alarm rate, and timeliness) of the non traditional surveillance data would be measured.

The gold standard surveillance data used in this research are described in this section are:

- ICD-9-coded <u>hospital</u> discharge diagnoses
- ICD-9-coded <u>outpatient</u> billing diagnoses
- Influenza surveillance data
- Hospital charts

2.1 ICD-9 Hospital Discharge Diagnoses

Many state health departments (approximately 60%) compile diagnoses of patients discharged from hospitals into datasets. In these hospital discharge datasets, discharge diagnoses are encoded using ICD-9-CM. The datasets also include dates of admission and discharge, home zip code, hospital zip code, and patient age. Because these datasets include diagnoses and locations, it is possible to construct reference epidemiological curves for zip codes, cities, counties, or states for diseases such as Influenza, Salmonella, respiratory syncytial virus, and rotavirus.

The following hospital discharge datasets were obtained for use in the research:

Pennsylvania Health Care Cost Containment Council (PHC4) dataset. Pennsylvania Department of Health provided a hospital discharge dataset that comprised all discharges for all hospitals in Pennsylvania for a three-year period.

Utah Hospital Discharge Database. The Utah Department of Health provided a discharge dataset for the Salt Lake region extracted from the Utah Hospital Discharge Database, which includes all discharges for all hospitals in Utah except Intermountain Shriners Hospital, which is exempt from reporting requirements because it is a charity hospital. Because that hospital specializes in Orthopedics, its absence is unlikely to be of significance in studies related to detection of outbreaks of infectious disease.

Indianapolis Discharge dataset. The Indianapolis Network for Patient Care provided a hospital discharge dataset for the Indianapolis region for the study period, covering 95% of hospital discharges in the region.

As an example of how these hospital discharge diagnosis datasets were used, the Mellon Pitt team defined four sets of ICD-9-CM codes: Pneumonia & Influenza (P&I), bronchiolitis due to Respiratory Syncytial Virus (RSV), Rotavirus gastroenteritis, and pediatric gastroenteritis due to all causes [5]. These sets of codes were used to aggregate cases from the discharge datasets to create weekly counts of these illnesses for the research.

2.2 ICD-9 Coded Outpatient Billing Diagnoses

Outpatient billing datasets are compiled from physician outpatient bills that are submitted electronically to insurers. These datasets are important for research in public health surveillance because a majority of patient interactions with the healthcare system for both acute and chronic disease occur as visits to physician outpatient offices, rather than hospitals; therefore, they potentially can be used to identify smaller outbreaks or outbreaks of diseases that do not often cause hospitalization. In addition to high-resolution due to volume and completeness of disease monitoring, these data have relatively low time latency, which is not of direct relevance to their use in this research as a gold standard, but is relevant both to their future use as gold standard surveillance data in prospective trials of early warning biosurveillance systems, as well as for use as a "non traditional" data source itself.

Surveillance Data, Inc (SDI) dataset. SDI receives claims from over 130,000 physicians across the U.S. on a daily basis. These physicians are using several common third-party carriers as intermediaries to transmit their claims, resulting in a relatively stable sample on an ongoing basis. As claims are sent, they are de-identified pursuant to HIPAA, transmitted to SDI, and loaded into a data warehouse by SDI.

Claims variables include date of service, patient and physician location at the five-digit zip code level, patient age and gender, all diagnoses for the visit by ICD-9 code and all procedures performed by CPT-4 code. Although data are available for all diagnoses, specific diagnostic code sets for respiratory and gastrointestinal illness typically were used in Bio-ALIRT to provide aggregated illness levels of clinically relevant diseases for specific pre-defined geographic areas. Although the source data contains specific street address information for both the physician office and the patient home that potentially could be available for public health operations purposes, data is available to SDI pursuant to HIPAA only at the five-digit zip code level for patients.

2.3 Influenza Surveillance Data

Influenza surveillance data are readily available from the Centers for Disease Control (CDC) website at http://www.cdc.gov/ncidod/diseases/flu/. They include:

1. Weekly state epidemiologist reports of influenza activity.
2. The weekly regional number of influenza-like-illness[1] (ILI) cases from the U.S. Influenza Sentinel Physicians Surveillance Network
3. The weekly regional number of positive influenza tests

A second source of Influenza data are state health department estimates. State health departments report the *estimated* level of influenza activity (based on item number two in the preceding list) in their states each week using four levels:

- *No Activity*: No cases of influenza or influenza-like illnesses (ILI) reported.
- *Sporadic*: Cases of influenza or ILI are reported, but reports of outbreaks in places such as schools, nursing homes, and other institutional settings have not been received.

[1] The CDC defines influenza like illness as fever (temperature of >100°F) plus either a cough or a sore throat.

- *Regional*: Outbreaks of influenza or ILI are occurring in geographic areas containing less that 50% of the state's population. A geographic area could be a city, county, or district.
- *Widespread:* Outbreaks are occurring in geographic areas representing more than 50% of the state's population.

(N.B. An additional source of influenza test results and influenza-like activity would be a program such as FluSTAR. FluSTAR is a network of 400 primary care physicians across the U.S. who have agreed to perform surveillance for influenza by tracking patients presenting with flu-like symptoms and testing a subset of these patients using a rapid point-of-care assay. This program provides collected behavioral and symptomatic information on patients presenting from the community with flu-like symptoms, as well as confirmation via the test result on the presence or absence of flu.)

2.4 Hospital Charts

Several studies employed gold standards derived by manual review of hospital charts. For example, Chapman et al. [6] reviewed a random sample of 1500 emergency room charts and categorized each visit as belonging to one of seven syndromic categories (e.g., respiratory, neurological, and gastrointestinal). They used these classifications as a gold standard against which to measure the accuracy of algorithms that analyzed free-text chief complaints (described in detail in Section 3.3).

3 LEADING INDICATORS

In this section, we discuss "leading indicators"--those data considered most promising by the authors of this report. These types of data were heavily studied and in some cases all four of the teams conducted research on a type of data.

For each leading indicator, we provide a subsection entitled "data characteristics" in which we discuss basic characteristics of the data that are relevant to their use in early warning biosurveillance. These characteristics include availability, latency, spatial granularity, and temporal granularity. We then provide a section that summarizes the available evidence about detectability developed by the BioALIRT program, but also drawing on prior work as warranted. In particular; we strive to describe the state-of-the-art of knowledge about the types of outbreak (e.g., organism, size) that might be detected by routine statistical monitoring of the indicator data. If relevant or available, we also describe work done to understand how to group raw data into analytic categories prior to further processing by detection algorithms. We also discuss results about known confounders and covariates (e.g., day of week effects) that can be potentially modeled to remove noise from the data.

3.1 Sales of OTC Healthcare Products

It is known that sick individuals treat their symptoms with over-the-counter medications, oftentimes before seeking medical care or in place of medical care and this is the motivation for investigating sales of over-the-counter (OTC) healthcare products as an early indicator.

A research literature review conducted by BioALIRT found two types of studies of OTC healthcare products. First, surveys of individuals who have had upper respiratory infections [7, 8] show that the majority of respondents recall self treatment with OTC healthcare products prior to seeking medical attention. Importantly, only a minority ever sought medical attention at all. Unfortunately these surveys did not ascertain (1) whether a purchase occurred versus use of a product that was already in the medicine cabinet at home, and (2) the day of illness on which purchases were made, leaving uncertainty as to how many unit sales might occur per sick individual and when these purchases might occur relative to the onset of symptoms.

Second, the literature review found six research publications that described retrospective studies of the sales of OTC healthcare products during known outbreaks [1, 9-13]. Those studies found strong and oftentimes early correlations between sales of cough and cold medications or sales of diarrhea remedies and outbreaks of disease Several found that rises in sales of OTC healthcare products preceded detection by public health authorities [9-13].

3.1.1 Data Characteristics

OTC healthcare products are sold by pharmacies, grocers, mass stores, and convenience stores. The manufacturers of these products imprint standard codes on the products and packages that can be scanned by optical scanners. In the United States and Canada, the Universal Product Codes (UPC) numbering system is used whereas in Europe the European Article Numbering (EAN) system is used [14]. Importantly, BioALIRT research discovered that 70% of these stores belong to 20 large national chains, that collect daily sales from their stores into national data warehouses [14]. The implication of this discovery is that 70% of all such sales nationally in the U.S. could be monitored in near real time—meaning with less than a 24 hour delay--by enlisting the assistance of fewer than 20 large national corporations. At the present time, a large project funded by the Commonwealth of Pennsylvania, the Alfred P. Sloan Foundation, and several state health departments has enlisted participation of eleven such chains accounting for 22,000 stores and roughly 50% of market share [14, 15]. The data are provided by store, so the spatial granularity of the data is roughly by "neighborhood," which can be represented by zip code of store, or longitude and latitude of store.

The UPC coding system is very specific, assigning unique numbers for each package from each manufacturer. So, for example, different and unique UPC codes exist for each size and flavor of each brand of cough syrup. As a result, approximately 7500 unique UPC codes exist for products that are used for self-treatment of symptoms of infectious disease. Obviously bottle size and flavor are not important distinctions in biosurveillance work; therefore, how to aggregate UPC codes into broader analytic categories such as *cough syrups, pediatric* is an important question also addressed by the research described below.

3.1.2 Experiments

The teams conducted the following types of studies of OTC sales data:

- Studies using seasonal disease activity (mixed disease outbreaks)
- Filtered correlation analyses
- Studies of purchasing behaviors
- Studies of multiple data streams
- Field testing

[N.B. the experimental methods used to evaluate OTC sales were also used to study other types of non traditional data. To avoid repeating the descriptions of methods in this report, we will describe a method in detail in a box the first time it is referenced. We will then refer the reader back to the box when it the method referenced in other sections.]

3.1.2.1 Studies using seasonal disease activity (mixed disease outbreaks)

Due to the rarity of "pure" outbreaks (i.e., outbreaks caused by one organism), we studied the effect of annual winter outbreaks of diarrhea and respiratory diseases on OTC sales data. These outbreaks typically involve organisms such as Rotavirus, Respiratory Syncytial Virus, Influenza virus, and Adenovirus. These studies used two basic methods that are described in Boxes 3.1 and 3.2 below: *Correlation Analysis*, and *Detection Algorithm Analysis*.

Box 3.1 Correlation Analysis

> <u>Correlation Analysis</u> measures signal strength and timeliness by using the correlation function to compare the indicator signal with a gold standard signal. The correlation function finds the time lag at which the correlation between two signals is maximized. A limitation of this method is that it typically finds the time lag at which the peaks of the two signals are most highly correlated, whereas for biosurveillance the time latency of the initial upticks of the signals are of most interest.
>
> The idea of studying corrrelation in the context of evaluation of public health surveillance data was first demonstrated in 1979 by Welliver and colleagues [1],

Box 3.2. Detection Algorithm Analysis

> Detection algorithm analysis attempts to address the limitation of correlation analysis by determining the date that a statistical detection algorithm first fires an alarm on non traditional data. It then takes the difference between this date and a gold standard reference date which may be externally determined by expert review, or may be the date when the same detection algorithm fires on gold standard surveillance data.

3.1.2.1.1 Pediatric Electrolytes

Pediatric electrolyte products are solutions of salts and water that are indicated for the treatment of dehydration in children ages five and under. As part of Bio-ALIRT, Hogan and colleagues studied the sales of electrolyte products during annual winter outbreaks in children ages five and under. The study involved six urban regions in three states over the course of three winters [5].

The researchers obtained a historical dataset of OTC sales data courtesy of Information Research, Incorporated. The dataset contained weekly counts of electrolyte products for regions in three states, representing nearly 100% of all sales. The gold standard surveillance data were hospital ICD-9 discharge diagnoses. The researches grouped hospital ICD-9 diagnoses into respiratory and diarrheal groups and created weekly counts to use as an indicator of actual disease activity in the large cities in those states against which to compare the sales of OTC products (Figure 3.1). There were six such cities studied and both correlation and detection algorithm analyses were performed for each city.

There were eighteen winter respiratory/diarrhea seasons to study (three winters in each of the six cities). Each of these 18 was considered to be an outbreak, so there were a total of 18 outbreaks to study.

The correlation analysis showed a consistently high correlation between sales of electrolyte products and hospital diagnoses of respiratory and gastrointestinal illness, with a mean of 0.90 (95% CI 0.87-0.93). The relative timing between sales and diagnoses ranged from an OTC lead of 8 weeks to an OTC lag of one week. On average, sales of electrolyte products preceded hospital diagnoses by 1.7 weeks (95% CI 0.5-2.9).

The detection algorithm analysis used the EWMA algorithm. The reference date of detection was established by EWMA analysis of the gold standard hospital diagnosis data. The study found that for the majority of these outbreaks that the up tick in sales preceded the up tick in hospitalization, but there was variation in the timing. The authors estimated, using regression analysis, that 25% of the variation was explained by differences in signal strength (for some winters and cities there were fewer or greater numbers of hospitalizations than sales). They also hypothesized that nosocomial outbreaks may contribute significantly to the unexplained variability and called for additional studies using outpatient diagnoses to control for this possible confounder.

This study added to the existing knowledge about OTC sales for outbreak detection in several ways: (1) it identified pediatric electrolytes as an important monitoring category, (2) it demonstrated the use of hospital discharge datasets as a gold standard, (3) it was the first study to include a sufficient number of outbreak to highlight the variability from year to year and city to city, and (4) it was the first study to include sufficient number of outbreaks to compute a confidence interval on the measurement of timeliness of detection.

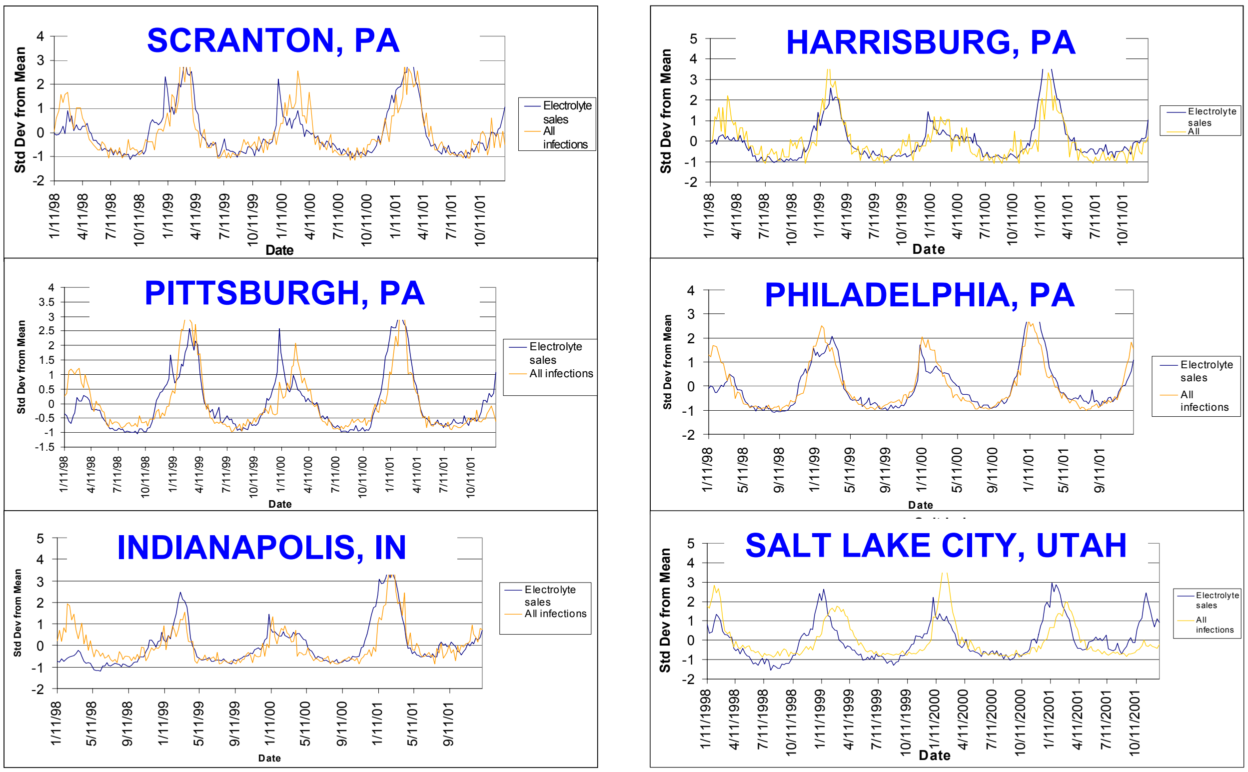

Figure 3.1. Correlation of weekly sales of electrolyte products and hospital diagnoses of respiratory or diarrheal illness in children age five and under. Data courtesy IRI, Utah DOH, Indianapolis Network for Patient Care, and PA HC4 Council

3.1.2.1.2 Flu Meds and Chest Rubs

Flu Meds and Chest Rubs are products used for the self-treatment of respiratory conditions. As part of Bio-ALIRT, Magruder used correlation analysis (Box 3.1) to demonstrate strong correlations between daily sales of flu medications and daily counts of physician's office diagnoses for acute respiratory conditions [16]. They also observed a correlation between sales of chest rubs and diagnoses of bronchitis. These similarities were noted when the time series were corrected for day-of-week effects, and when each was normalized by its respective average count. In the case of flu meds, correlations were calculated between sales of flu medications and corresponding clinical data at multiple lags, for six different regions within the states of Maryland, Virginia, and the District of Columbia. These correlations were consistently high, with a mean value (at the time lag corresponding to the peak correlation) of 0.895. The relative timing between OTC sales and physician office visits, as measured by peak correlation, ranged between an OTC lead of seven days, and an OTC lag of two days. On average, the OTC flu meds were observed to anticipate physician office ILI diagnoses by 2.8 days.

With the exception of Welliver et al [1], the analysis of Magruder was the first to provide quantitative estimates of the lead time of flu medications over clinical sources. Welliver observed a strong peak in cold remedy sales just at the beginning of a rise in encounters with clinical patients known to be infected with influenza B virus, and one week before the peak in those encounters. He also observed an earlier rise in cold remedy sales approximately coincident with the early winter rise in non–influenza respiratory virus activity. This early study was based on these two outbreaks in a single city, and utilized OTC data limited to a time resolution of one week.

3.1.2.1.3 Cough, Cold, Sinus, and Allergy Medications

Products in the cough, cold, sinus, and allergy categories are used for self-treatment of respiratory conditions. As part of Bio-ALIRT, Campbell and colleagues [17] used the detection algorithm method (Box 3.2) to assess the lead time of sales of these categories of OTC medications with respect to physician office visits. The study examined ten major US cities for a period of three years (2000-2002). The OTC sales data were aggregated weekly for each of eight categories (adult and pediatric versions of cough, cold, sinus, and allergy). The physician office visit data contained 22.5 million anonymized records corresponding to visits with respiratory-related ICD9 codes for the ten cities during the 2000-2002 period.

The beginning of the seasonal rise of respiratory ICD9 codes was identified by inspection for each year for each city, giving a total of 30 "outbreaks." A number of simple detectors (e.g., AR) were applied to both the OTC and office visit data for the periods preceding and following the outbreaks. For a number of categories (with the notable exception of allergy medications), OTC detection led office visit detection over a plausible range of false alarms (.5-1 false alarms per month per city). The lead time for the best categories (e.g., sinus, pediatric cough) was 1-7 days over this false alarm range (Figure 3.2)

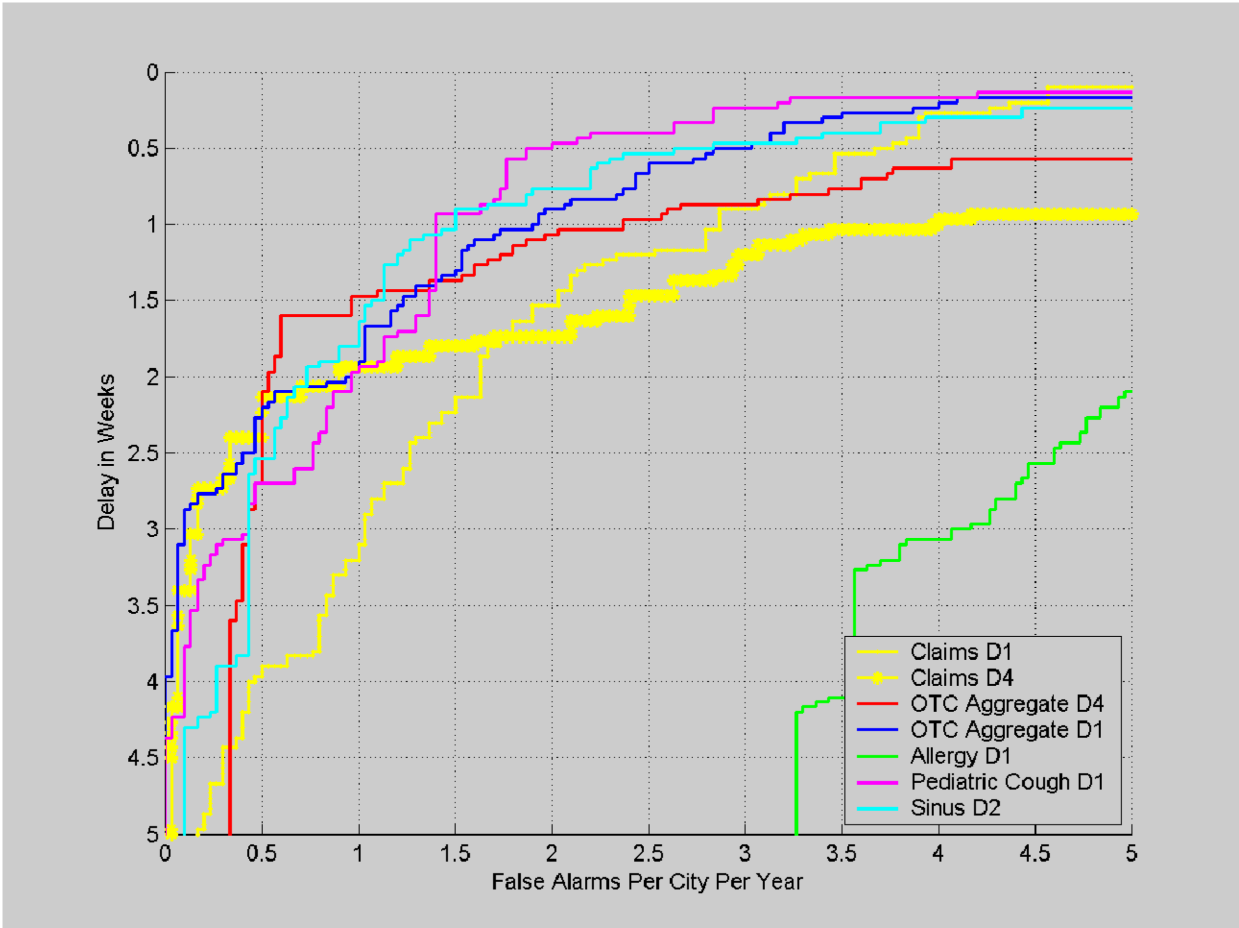

Figure 3.2 AMOC curve showing the results of Detection Algorithm Analysis of OTC and office visit data for seasonal respiratory outbreaks. For a number of categories (with the notable exception of allergy medications), OTC detection led office visit detection over a plausible range of false alarms.

3.1.2.2 Correlation with pure outbreaks

As summarized above, research published prior to BioALIRT found correlations between OTC sales and actual outbreaks of influenza and cryptosporidiosis [9-12]. The advantage of studying pure outbreaks is that a source of variability is removed and furthermore the problem of removing through modeling any background disease activity is eliminated.

Because of the rarity of pure outbreaks, no Bio-ALIRT research was done with pure outbreaks.

3.1.2.3 Filtered correlation

Related to trying to reap the advantages of studying pure outbreaks, is the technique of filtered correlation. This technique has been applied in business analysis [18]. For the general theory, see [19].

Filtered correlation refers to a technique where a filter is applied to the raw data before the correlation analysis is conducted. The reason for this filtering is to remove seasonal trends that may dominate the correlation. That is, when an unfiltered (raw) time series of OTC sales are correlated with unfiltered time series of clinical visits, the correlations, and in particular the time-lag that maximizes the correlation, is sometimes determined mostly by the strong annual cycle present in both these types of data

As part of Bio-ALIRT, several researchers investigated how the time lead, as measured by correlation analysis, varies with how the raw data are filtered to emphasize different time-scales [20]. These researchers have observed that the apparent time lead of OTC sales relative to clinical events is indeed dependent on the time-scales of variation that are passed by the filter. For example, in [20], OTC flu med sales and acute respiratory physician's office diagnoses are separated into an annual cycle and a residual (non-annual, and non-weekly) variation. In a study of six regions in the Virginia-Maryland-DC area, it was found that the annual cycle of OTC flu-med sales precedes the corresponding annual cycle of physician office visits by two to 16 days, while the residual, non-annual variations of OTC lead by minus eight to three days, as measured by peak correlation lags. They speculated that drugstore promotions, consumer habits, and multiple types of illness may all be contributing to the observed lead-time, and that these factors may vary on different time scales.

Bloom and Cheng (unpublished results) have gone a step further and divided time series into a variety of time-scale and time-of-year segments (called "epochs"), measuring the correlation time lead in each epoch. This approach provides a view of the dependence of correlation time lead both on time-scale of variation and on the epoch (approximately 1-month long period) in which the time-lead is measured. In this work, spectral analysis (see for example [21]) is employed. The variation of the time series under study is separated into a portion that can be formed by a sum of long-wavelength sinusoids (the "low frequency component"), a portion that can be formed by a sum of medium-wavelength sinusoids (the "band-passed" component) and a portion formed as a sum of shorter-wavelength sinusoids (the "high frequency" component).

Studying military physician diagnoses and military pharmacy data from the Norfolk, Virginia area, they found that peak cross-correlations occur at different lags depending on whether or not the low frequency component has been included. In pairs of daily data streams exhibiting strong 7-day cycles, peak cross-correlations in the high frequency component almost invariably were found at zero lag, even though the 7-day cycle and its harmonics were filtered out. On the other hand, in a band-passed component, admitting wavelengths roughly shorter than 30 days and longer than 7 days, peak positive cross-correlations were often found at lags other than zero. The appearance of effects at zero lag in the high-passed data part of the spectrum driven by day-of-week influences after 7-day cycles have been removed may be attributable to same day-of-week influences that occur at random intervals as opposed to regular intervals. These zero lag effects were not entirely removed by pair-wise deletion of holidays and other simultaneous "exceptional" days. It was also found that the magnitude of the peak cross correlation between certain pairs of diagnostic and over-the-counter data exhibited a tendency to increase during the winter season.
Finally, it was observed that these types of results showed significant variation for identical epochs when time-series components were formed giving different emphasis to different wavelength sinusoids.

3.1.2.4 Studies of purchasing behaviors

The importance of surveying sick individuals to determine early illness (preclinical) behavior was recognized by the researchers during the second year of the program and several such studies were designed and initiated (and are in progress) Such studies can be highly complementary to other studies, being informative about timing and strength of a signal that can be expected in different outbreak situations. As part of Bio-ALIRT, WRAIR is in the process of conducting as survey of adult patients and the University of Pittsburgh is conducting a survey in pediatric patients.

3.1.2.5 OTC Analytic categories

There are approximately 7500 different UPC-coded OTC healthcare products that are indicated for symptomatic treatment of infectious disease. This list of products can be viewed as thousands of possible surveillance data channels that may be highly redundant or uninformative; thus, it is important to consider ways to aggregate these products before using them for surveillance. The clustering not only strengthens statistical stability of the total counts, but also provides an opportunity to control some of the effects of sales promotions which attempt in general to induce buyers to select one UPC coded product to another within the same general category like "cough syrups."

As part of Bio-ALIRT, Magruder and colleagues developed a method for forming the required clusters and correcting for promotions that is described in Box 3.3 [20]. The method includes positioning products in a taxonomy based on simple descriptive product information, followed by empirical cluster analysis based on observed sales histories. The results of the empirical clustering followed expectations in some respects, but also uncovered some less predictable associations between products. For example, it was discovered that sales patterns of pain-relieving pills were strongly similar to allergy medicine sales in the National Capital Area.

Box 3.3 Empirical Clustering Algorithms

Empirical clustering algorithms automatically find groupings of indicator signals. They use a measure of similarity between different signals, based upon historical temporal patterns. Clusters are formed iteratively. In the beginning, every indicator is treated as a separate group. Then, at each stage, the two most-similar groups are combined, reducing the number of separate groups by 1. This continues until a manageable number of groups remain. The strength of the empirical clustering approach is that relationships between different indicators can be judged on a quantitative basis, including relationships that might not have been suspected. The limitation is that sometimes these connections are coincidental, and thus, not repeatable.

3.1.2.6 Research on covariates

As part of Bio-ALIRT, Florio [22] has evaluated a number of different covariates that may explain observed day-to-day variability in sales, including, day-of-week, linear time trend, seasonal time trend, outdoor temperature, and promotion indicators. She quantitatively measured the contribution of each to "explain" (in a regression model sense) variations in sales of OTC medications. She found that all of these factors can be useful explanatory variables for some types of medications, while others can be modeled just as well by a subset of the factors.

3.1.2.7 Analyses using semi-synthetic data

Due to the rarity of actual outbreaks, researchers are increasingly injecting artificial (synthetic) "spikes" (e.g., triangular shaped increases in daily counts) into real data. A limitation of this approach is that assumptions must be made about the expected effect of an outbreak on non traditional data, the very unanswered question that the BioALIRT research on non traditional data is attempting to address. For this reason, the technique mainly has applicability to comparison studies of algorithms (e.g., how small of a spike could detection algorithm A detect). This method was pioneered by Goldenberg et al in a study of detectability of Anthrax from OTC sales data [23]

As part of Bio-ALIRT, Wallstrom [24] developed a method to use actual surveillance data collected during known outbreaks to derive the shape and scale of the injected signal. An example analysis used injected signals to evaluate the ability of an algorithm to detect a cryptosporidium outbreak.

Box 3.4 Semi-synthetic Injection Methods

> Semisynthetic injection refers to the practice of injecting geometrical shapes of various sizes into actual surveillance data for the purpose of studying detectability of different size and shape perturbations in the data that might result from outbreaks. The resulting datasets then contains an artificial outbreak effect that can be manipulated as to day of onset, duration, size, and shape. Detection algorithms can then be run to determine when the outbreak would be detected relative to the start of the injection. From multiple studies, ROC or AMOC curves can be drawn. Some of the assumptions underlying semisynthetic injection can be addressed by conducting sensitivity analyses over the shapes and heights of the injected data. However, the method can only produce results about the size of the data perturbation that can be detected, not about the size of the actual outbreak that can be detected.

3.1.2.8 Field Testing

The National Retail Data Monitor collects "case reports" from public health users of the systems about outbreaks that have been detected or missed by routine monitoring of OTC sales data. To date, five case reports relevant to OTC monitoring have been compiled and can be accessed by authorized public health officials at https://www.rods.pitt.edu/cases/. [N.B. the 's' in *https*]. They are:

- Case Study #2: Utah 2003 Flu (html, PowerPoint)
- Case Study #3: Southern California Wildfires 2003 (html, PowerPoint)
- Case Study #4: Philadelphia-Gastroenteritis (Norovirus?) at Loews Hotel (html, PowerPoint)

- Case Study #5: Philadelphia Shigella(?) Outbreak 2003 (html, PowerPoint) (currently under review)
- Case Study #6: Fort Wayne, Indiana, Cough&Cold, 2003 (html, PowerPoint)

These case studies add to the laboratory evaluations of historical data the following: Influenza has a strong early effect on sales of OTC cough and cold, pediatric antipyretics. Air pollution due to smoke causes marked effects on sales of bronchial remedies. The results for gastrointestinal outbreaks have been negative for relatively small, protracted outbreaks of norovirus and Shigella.

3.1.3 Conclusions about Timeliness and Detectability

Sales data are highly available in near real time. The following table summarizes the results about OTC categories/diseases.

	Positive	Negative	Notes/caveats	References
Pediatric electrolytes	Can detect winter diarrheal/respiratory in children ages 5 and under earlier than hospital data		Variability of effect noted from year to year and city	Hogan 2003
Flu meds	Acute respiratory detected earlier than office diagnoses			Magruder 2003
Chest rubs	Acute respiratory detected earlier than office diagnoses			Magruder 2003
Diarrheal remedies	Can detect large cryptosporidium outbreaks 3-4 weeks before conventional methods		Literature review	Five studies found on literature review
Pediatric cough	Winter respiratory			Campbell
Cold	Winter respiratory			Campbell
Sinus	Winter respiratory/ bronchial diagnoses			Campbell
Allergy		Winter respiratory (later than office records)		Campbell
Bronchial remedies	California wildfire showed effect on day that air quality deteriorated			Case study

3.2 Advice Lines and Appointment Calls

Hospitals and other healthcare facilities may operate **advice lines** and **appointment centers** that use computerized record keeping and therefore collect data that are of potential value for early detection of a covert biological event.

Advice lines are facilities such as nurse call centers that receive telephone calls from persons requiring information, triage, or immediate assistance. Government agencies, healthcare systems (especially HMOs), or private contractors may operate these facilities. Advice lines gather information about a caller's complaints that enables them to allocate appropriate clinical resources to that patient with appropriate urgency and importantly they doc. This practice results in an electronic record carrying syndromic information.

In some cases, medical insurance policies stipulate that covered members call the advice lines before they see a physician. In theory, there can be zero latency between the time of the call and the time at which relevant information is available to public health offices. It has been observed that, for syndromes of interest, the nurse hotline calls outnumber physician office visits by a factor of *** [25] .

Calls for appointments with physicians are recorded electronically by some health maintenance organizations (HMOs) along with chief-complaint level information and a patient ID that can be linked to other patient information such as age, gender, and home and work address. At (at least) one HMO, appointment services and an advice line are integrated into a single service. In this case, a large fraction of appointment calls related to acute illnesses are also referred to an advice nurse [25].

Wagner et al discusses advice lines and appointment call centers in more detail [2].Additional research in this area has been conducted by Platt and colleagues at the Harvard Pilgrim Healthplan [26].

3.2.1 Data characteristics

Telephone advice and appointment data typically include the caller's identity and medical record or insurance number (although all such identifiers were removed from the data sets used in this research). They also include timestamps, location, reason for call, and disposition. They may include a categorization of the problem into a finite number of pre-specified diagnostic or management categories (e.g., practice guidelines used by the nurse to manage the call). This categorization is made by a health professional—typically a registered nurse—based on information provided by the caller.

Box 3.5 Linked Record Analyses

> The linked-patient method relies on the ability to compare two indicators or an indicator and gold standard on a patient-by-patient basis. With this method, the time of a patient encounter in one data source can be compared to the time of an encounter of the same patient in a different data source, to measure relative latency of the two data sources. Classification of illness categories can also be compared between the two data sources by measuring the frequency with which a specific illness classification in one source results in the same or another classification for the same patient in a different source.

3.2.2 Experiments

As part of Bio-ALIRT, Espino and colleagues used correlation analysis (See Box 3.1) to evaluate telephone triage (TT)—a type of advice line. They studied emergency room TT data and after hours TT data from a commercial triage service company serving two states. In emergency telephone triage, an insured person calls the TT center to obtain authorization to visit an emergency room. In after hours TT, the insured person calls the TT center instead of the their physician

The dataset included, in addition to the verbatim complaint, the practice guideline selected by the registered nurse to manage the patient call. Using influenza surveillance data from the Centers for Disease Control as a gold standard, they measured the timeliness of increased numbers of patients managed using guidelines from a set of guidelines that they considered to be respiratory. They found that calls to emergency room TT for respiratory problems increased one to five weeks ahead of increases in influenza surveillance data collected by the CDC. [27]

As part of Bio-ALIRT, Henry and colleagues [25] studied a one-year dataset from an HMO in the National Capital Region that included nurse advice calls encoded according to a dictionary of advice protocols (similar to practice guidelines), using ICD-9 coded physician encounter data as a gold standard. This linked analysis (see Box 3.5) was done behind the HMO "firewall," so that patient IDs could be used to link the data sources without violating privacy regulations. Time-lag distributions were measured between nurse advice calls resulting in a physician appointment and the actual physician encounter, and mappings could be constructed, showing the fraction of patients within a given nurse advice syndrome group that eventually were diagnosed into a given physician ICD9 syndrome group.

Henry et al found that 60% of calls classified by the nurse in an acute respiratory syndrome group were eventually diagnosed by the physician in an acute respiratory syndrome grouping for ICD9 codes. Also, the median time lag between the nurse advice call and the corresponding physician encounter was 5 hours. 32% of calls classified as gastrointestinal resulted in gastrointestinal physician primary diagnoses, and these had a median lag of 4 hours. Other categories showed weaker correspondences between the nurse advice protocols and the physician diagnoses. It is not known yet to what extent this might be due to the way the protocols and diagnoses were aggregated. That is a subject for further research.

3.2.3 Conclusions about Timeliness and Detectability

The following table summarizes the results about advice lines and appointment call lines.

	Evidence for	Evidence against	Notes/caveats	References
After hours calls for which nurse used respiratory practice guideline		little effect seen		Espino
ED authorization calls for which nurse used respiratory practice guideline	calls to emergency room TT for respiratory problems increased one to five weeks ahead of increases in influenza surveillance data collected by the CDC			Espino
HMO triage calls for acute respiratory syndrome	Linked analysis showed that 60% of those individuals who were classified as acute respiratory and who eventually saw physicians actually had ICD-9 respiratory diagnosis. Time advantage for this HMO was 5 hours over office visit.		Some % of patients did not seek further care at the HMO and therefore would not have shown up in any other type of data except the call data.	Henry
HMO triage calls for acute gastrointestinal syndrome	32% of calls classified as gastrointestinal resulted in gastrointestinal physician primary diagnoses, and these had a median lag of 4 hours			Henry
HMO triage calls for other syndromes		Weaker Hemorrhagic, Dermatological, Fever, and Neurologic.		Henry

3.3 *Free-text Chief Complaints*

Chief complaints are short phrases describing one or more medical problems with which a patient presents to a medical facility. A chief complaint may be a direct verbatim transcription using the patient's words, or an interpretation by a nurse, and it is entered into the registration computer system by a triage nurse or registration clerk at the time of registration to an emergency room or clinic. Chief complaints are also encoded into ICD-9 codes by some hospitals. This step usually occurs, however, several days after the visit and is the subject of a later section.

3.3.1 Data Characteristics

Unlike other data described in this report, chief complaints are uncoded data; that is, they are recorded in English and as a result contain significant lexical variability in the form of synonymy, abbreviations (such as "n/v/d" for "nausea, vomiting, diarrhea"), misspellings, and truncations. For this reason, chief complaints must be processed by natural language processing (NLP) techniques and part of the work performed in the BioALIRT program in this area measured the accuracy at which this translation could be accomplished.

Chief complaints are highly available surveillance data. Due to computerization of ED registration and HL-7 infrastructure in healthcare, chief complaints are available from most health systems in the country immediately after a patient registers at an emergency care facility. They therefore provide early clinical information. The data are identified at the level of the individual patient and can therefore be linked to home location and analyzed spatially with data from other patients to detect clusters of cases. Chief complaints from outpatient visits are less available electronically due to lesser penetration of computer systems into office practice.

3.3.2 Experiments

3.3.2.1 Mellon Pitt Experiments

As part of Bio-ALIRT, Mellon-Pitt conduced many experiments that fell into two types: (1) Measuring the classification accuracy of various natural language processing algorithms that classified patients into syndromic categories (Box 3.7), and (2) measuring the ability to detect outbreaks using time series constructed from the output of the natural language processors using the Detection Algorithm Method as described in Box 3.2.

Box 3.7 Methods for Measuring Accuracy of Case Detection

> To evaluate algorithms such as free-text classifiers that analyze data about <u>one</u> person, we used standard approach to measure the accuracy of case classifiers and produce as measures sensitivity and specificity of classification. This type of experiment requires a set of individuals known to have the condition and a set known not to have the condition to serve as the gold standard. In addition to sensitivity and specificity measurements, ROC curve analysis is typically performed with reporting of the area under the ROC curve as a measure of classification accuracy.
>
> Although accuracy of case detection is not a direct measure of outbreak detection, it provides information about the expected signal-to-noise ratio in aggregate data. Additionally, studies of case accuracy are much easier to perform than studies of outbreak detection because there are many more cases of respiratory illness in a population during the course of a year (typically thousands in an urban region) than there are outbreaks of respiratory illness (typically one).

Case detection from chief complaints. Table 3.1 summarizes results of investigations of case-detection accuracy for seven bioterrorism prodromes (prodromes are early presentations of disease, more commonly referred to as “syndromes” in the bioterrorism detection literature). The largest and most recent of the studies involved thirteen years of patient admissions to an ED in Pittsburgh, including over 500,000 cases. The evaluations summarized by the chart differ slightly methodologically in the type of gold standard (e.g., ICD9 discharge diagnoses or manual review of patient charts) but are otherwise comparable.

The stronger gold standard is patient chart review. Those studies have shown that automatically processed chief complaints can detect the majority of patients with fever and with respiratory and gastrointestinal syndromes. Moreover, the sensitivity of case detection from free-text chief complaints is higher than that from ICD9 codes (which also are available only with up to several days of time delay). Compared with manual classification of patients based on their ED reports, automatic classification and extraction from chief complaints detected patients with acute lower respiratory illness with a sensitivity of 0.77 and a specificity of 0.90, patients with acute infectious gastrointestinal illness with a sensitivity of 0.63 and specificity of 0.92, and febrile illness with a sensitivity of 0.61 and a specificity of 1.0. These findings indicate that the majority of patients with these syndromes have chief complaints related to the syndromes and that they can be identified automatically by the NLP techniques used. Evaluations using the weaker gold standard (discharge ICD9 codes) show similar results with sensitivities between 0.13 for botulinic syndrome and 0.74 for gastrointestinal. Research using ICD9 discharge codes as a gold standard is much less expensive and time consuming to conduct, so the fact that the results are comparable with the more labor-intensive methods are significant in that the easier gold standard may be sufficient for research and therefore will accelerate future research.

Table 3.1. Accuracy of Syndrome Detection from Chief Complaint Data

Classifier being tested	Gold Standard Cases	Sensitivity	Specificity	LR+ (AUC)
Respiratory Syndrome				
CoCo Respiratory *[28]*	UDOH respiratory with fever*	0.52 [0.51-0.54]	0.89 [0.89-0.90]	5.0 [4.74-5.22]
CoCo Respiratory [29]	Human review	0.77 [0.59-0.88]	0.90 [0.88-0.92]	7.9 [5.8-10.8]
CoCo Respiratory*[28]*	Utah ICD-9	0.60 [0.59-0.62]	0.94 [0.94-0.95]	10.5 [9.90-11.05]
CoCo Respiratory with Fever *[29]*	Human review	0.22 [0.06-0.55]	0.99 [0.98-0.99]	24.5 [5.7-105.3]
ICD-9 Respiratory *[30]*	Human review	0.44 [0.29-0.61]	0.97 [0.96-0.98]	15.6 [8.6-28.1]
CoCo Respiratory [31]	Pittsburgh ICD-9	0.36 [0.35-0.36]	0.96 [0.96-0.96]	9.0 [8.9-9.2]
Gastrointestinal (GI) Syndrome				
CoCo GI*[28]*	UDOH Gastroenteritis without blood	0.71 [0.69-0.74]	0.90 [0.90-0.90]	7.1 [6.80-7.51]
CoCo Acute Infectious GI [32]	Human review	0.63 [0.35-0.85]	0.94 [0.92-0.96]	12.2 [8.3-18.] (0.82 [0.75-0.90])
ICD-9 Acute Infectious GI [32]	Human review	0.32 [0.14-0.54]	0.99 [0.98-0.99]	37.1 [16.2-85.3]
CoCo GI*[28]*	Utah ICD-9	0.74 [0.72-0.76]	0.92 [0.92-0.92]	9.5 [9.04-9.94]
CoCo GI with diarrhea *[29]*	Human review	0.12 [0.06-0.22]	0.99 [0.99-0.99]	81.1 [17.56-374.4]
CoCo GI with vomiting *[29]*	Human review	0.16 [0.11-0.24]	0.99 [0.99-0.99]	105 [24.85-444.33]
CoCo GI [31]	Pittsburgh ICD-9	0.51 [0.50-0.52]	0.95 [0.95-0.95]	10.4 [10.2-10.5]
Neurologic/Encephalitic Syndrome				
CoCo Neurologic*[28]*	UDOH Meningitis / encephalitis	0.47 [0.32-0.63]	0.93 [0.93-0.94]	7.1 [4.98-9.99]
CoCo Neurologic*[28]*	Utah ICD-9	0.72 [0.69-0.76]	0.95 [0.94-0.95]	13.5 [12.57-14.41]
CoCo Neurologic [31]	Pittsburgh ICD-9	0.46 [0.46-0.47]	0.94 [0.94-0.94]	8.1 [7.9-8.3]
Rash Syndrome				
CoCo Rash*[28]*	UDOH Febrile illness with rash*	0.50 [0.40-0.59]	0.99 [0.99-0.99]	55.6 [44.25-69.91]
CoCo Rash*[28]*	Utah ICD-9	0.60 [0.52-0.67]	0.99 [0.99-0.99]	80.9 [67.43-97.07]
CoCo Rash [31]	Pittsburgh ICD-9	0.38 [0.36-0.39]	0.99 [0.99-0.99]	48.8 [46.1-51.6]
Botulinic				
CoCo Botulinic*[28]*	UDOH Botulism-like syndrome	0.17 [0.05-0.45]	0.998 [0.998-0.999]	104 [28.57-381.86]
CoCo Botulinic*[28]*	Utah ICD-9	0.22 [0.13-0.36]	0.999 [0.998-0.999]	167 [89.07-312.90]
CoCo Botulinic [31]	Pittsburgh ICD-9	0.14 [0.13-0.15]	0.99 [0.99-0.99]	24.4 [22.6-26.4]
Hemorrhagic				
CoCo Hemorrhagic [31]	Pittsburgh ICD-9	0.38 [0.37-0.38]	0.99 [0.99-0.99]	40 [38.6-41.3]
Constitutional Syndrome				
CoCo Constitutional [31]	Pittsburgh ICD-9	0.40 [0.40-0.41]	0.95 [0.95-0.95]	8.7 [8.5-8.9]
Fever				
Keyword Fever [33]	Human review	0.61 [0.51-0.69]	1.0 [0.96-1.0]	[2]
Fever from ED report [33]	Human review	0.98 [0.94-0.99]	0.89 [0.82-0.94]	9.3 [5.3-16.2]

*Classifier trained on less specific training classifications than gold standard: gold standard required documentation of fever in the patient record.

Outbreak detection from chief complaints. Ivanov [32] measured the timeliness, sensitivity, and specificity with which seasonal outbreaks of adult respiratory illness, pediatric respiratory illness, and pediatric gastrointestinal illness from free-text chief complaints. Table 3.2 shows that detection from chief complaint data is timelier than from ICD9 diagnoses.

---

[2] $LR+ = \frac{sensitivity}{1 - specificity} = \frac{0.61}{1-1}$

Table 3.2. Detection Algorithm Analysis of Timeliness of Detection from Chief Complaints

Syndrome	Gold Standard Outbreak	Sensitivity	Specificity	AUROC	Timeliness (days)
Respiratory	Seasonal outbreaks of adult respiratory illness (P&I)	100%	100%	1.0	11 days 95% CI (-10,33)
Respiratory	Seasonal outbreaks of pediatric respiratory illness (bronchiolitis, P&I)	100%	100%	1.0	10 days 95% CI (-15;35)
Gastrointestinal	Seasonal outbreaks of pediatric gastrointestinal illness (rotavirus gastroenteritis)	100%	100%	1.0	29 days 95% CI (4, 53)

Field Testing

The RODS Laboratory collects "case reports" from public health users of the systems about outbreaks that have been detected or missed by routine monitoring of chief complaint data. To date, five case reports relevant to chief complaint monitoring have been compiled and can be accessed by authorized public health officials at https://www.rods.pitt.edu/cases/. [N.B. the 's' in *https*]. They are:

- Case Study #1: PA Carbon Monoxide 2003 (html, PowerPoint)
- Case Study #2: Utah 2003 Flu (html, PowerPoint)
- Case Study #3: Southern California Wildfires 2003 (html, PowerPoint)
- Case Study #4: Philadelphia-Gastroenteritis (Norovirus?) at Loews Hotel (html, PowerPoint)
- Case Study #5: Philadelphia Shigella(?) Outbreak 2003 (html, PowerPoint) (currently under review)

These case studies add to the laboratory evaluations of historical data the following: Influenza has a strong early effect on free text chief complaints in the constitutional and respiratory categories. Air pollution and small carbon monoxide events have marked effects on chief complaints in the respiratory category. The results for gastrointestinal outbreaks have been negative for relatively small, protracted outbreaks of norovirus and Shigella.

3.3.2.2 JHAP Experiments

As part of Bio-ALIRT, A parser has been developed at JHUAPL known as the Chief Complaint Parser (CCP). It follows a two-step process: chief complaint records are first classified into more narrowly defined symptom categories, which are then combined into broader syndrome groups. The first step is performed using a method similar to the cosine vector method for determining document relevance in information retrieval; the second step is based on logical rules. The two-step approach parallels the way in which medical experts define syndrome groups: based on lists of individual symptoms or conditions – fever, wheezing, nausea, pneumonia – that are combined into groups using ANDs, ORs, exclusions, and other logical constraints.

The symptom-based approach has proven useful. It allows certain types of revisions to the syndrome group definitions – excluding a symptom that had formerly been included, adjusting an age constraint – to be implemented by simply reformulating the relevant logical rules. During the recent SARS outbreaks, the approach allowed SARS surveillance to be quickly implemented by constructing a new syndrome group out of logical combinations of existing symptom categories. Creating new symptom categories involves more effort; but once the categories are implemented, they can be used and reused as building blocks for syndrome groups as desired.

Close examination of the data shows that significant inter-hospital variation in chief complaint exists. Some sources favor more formal descriptions that resemble ICD-9 code text, while others are more colloquial. The quantity, form, and predominant meaning of abbreviated expressions, as well as the quantity and form of spelling errors, have been observed to differ between hospitals. Variation may be due to formal policy, informal convention, or the individual predilections of hospital staff. Tuning a system to perform well on one hospital's data does not necessarily guarantee the same level of performance on others. Although most syndrome groups – notably respiratory and gastrointestinal – appear to have a core of "easy" cases consisting of relatively commonly occurring conditions that are represented in the chief complaint as canonical text expressions, each has a percentage of cases expressed in more idiosyncratic and/or error-filled terms. The distributions of these "harder" cases – whether they are randomly distributed or whether they cluster in certain hospitals or symptomatology – will determine how important it is to classify them correctly for purposes of surveillance.

3.3.3 Conclusions about Timeliness and Detectability

Chief complaints are the earliest data available on most electronic hospital information systems for patients who physically present to hospitals and clinics. Early detection of cases of interest from chief complaints can be accomplished with sufficient accuracy to support detection of large seasonal outbreaks of respiratory and diarrheal illness in children and adults, as demonstrated by the results presented in Tables 3.1 and 3.2. A key question is the lower bound on size of outbreak that can be detected. Improving the accuracy of individual case detection by use of more data about the patient or better natural language processing will improve detectability, as will better algorithmic processing especially with attention to spatial and temporal analysis.

3.4 ICD-9 Coded Chief Complaints

The Health Care Financing Administration developed ICD9 codes to represent patient diagnosis for the purpose of processing Medicare claims. This coding system has been adopted more widely for all billing purposes. As a result, most encounters with physicians or other healthcare providers result in one or more ICD-9 coded diagnoses, assigned either directly by the healthcare provider or by professional coders. The Department of Defense (DoD) uses ICD9 codes in the Standard Ambulatory Record (SADR) at all military treatment facilities (MTFs).

3.4.1 Data Characteristics

There are more than 10,000 ICD9 codes available. In the DoD SADR database, ICD9 codes are entered for all diagnoses made during an outpatient clinic or emergency room visit and most of these codes are entered into the electronic system on the same day of care, although it may take up to three days to be transferred to the centralized database.

Although civilian outpatient clinics and emergency rooms ultimately encode diagnoses using ICD9 for billing, delays in coding that can be a month or more reduce their value for use in early warning systems. As discussed in the previous section, free-text chief complaints, captured at the time of registration an emergency care setting, can be processed using natural language parsers, providing an alternative to IDC9 coded diagnoses for generation of fast, timely reports on the health status of a particular community.

In the military treatment facilities, where active duty military as well as family members and retirees can seek care, the ICD-9 encoding is done by the physician, so timely availability of military outpatient ICD9 coded diagnoses makes them an attractive candidate for use in systems that are focused on early detection of disease events, including potential bio-events. Because these codes can be easily retrieved electronically, they have been successfully incorporated in military non-traditional disease surveillance activities like the ESSENCE program and MDSS. A potential downside to employing provider-entered codes is accuracy.

ICD9 codes must be aggregated into syndrome groups for surveillance of disease events. Early in the history of research in this field, assignment of particular codes to syndrome groups was based on the clinical opinion of medical experts specializing in infectious diseases, emergency medicine and preventive medicine. More recently, stepwise statistical tests combined with graphs of seasonal trends have been used to produce the best signal for detection. For example, a joint CDC, NYCDoH and DoD evaluation of ICD9 codes and syndrome definitions yielded an analysis on a subset of respiratory codes using pneumonia diagnoses as the gold standard. (Foster, Pavlin, and Marsden-Haug, unpublished data). Additional research about other syndrome groups that can potentially be created using ICD9 codes is needed.

3.4.2 Experiments

As part of Bio-ALIRT, several validation studies have been done to demonstrate the sensitivity of military ADM ICD9 codes by comparing them to a gold standard. Lee et al compared outpatient weekly percentage of respiratory syndrome group visits and an ILI sub-set in the National Capital area to the Southeastern region of the CDC's sentinel physician influenza surveillance program. Over two years, there was high correlation (0.81-0.89) between the civilian and military data (Lee, unpublished data)

Additionally, Marsden-Haug et al compared the ESSENCE Respiratory and ILI syndrome groups and associated regression-based detectors to the traditional, active surveillance system operated by the Naval Health Research Center (NHRC). The NHRC system has a nurse on site at each military treatment facility (MTF) to record the number of recruits who meet a case definition of Febrile Respiratory Illness and a threshold is used for alerts. The correlation of weekly Febrile Respiratory Illness case detection with Respiratory and ILI syndrome groups, from June 1998 to January 2002, was evaluated in nine MTFs. The MTFs with the strongest degree of correlation included: Fort Leonard Wood (r=0.8140) Fort Knox (r=0.7486), and Lackland Air Force Base (r=0.7006). (Marsden-Haug, unpublished data). Another validation study by Bentancourt et al compared ICD9 code-based syndrome groups assigned by ESSENCE with emergency room encounter diagnoses abstracted from chart reviews at three National Capital Areas military hospitals. Overall, accuracy of the ESSENCE ICD9 codes was high, with average sensitivity ranging 75 % to 93% and specificity ranging 95% to 96% for respiratory, gastrointestinal and fever syndrome groups (Bentancourt, Pavlin, Foster, Hakre, unpublished data).

Additional work in this area has been done by Espino and Tsui [30, 34] for detection of cases of respiratory and constitutional syndromes from ICD-9 coded chief complaints and for the detection of influenza outbreaks.

3.4.3 Conclusions about Timeliness and Detectability

Surveillance systems using electronic ICD9 code diagnoses can be effective tools for the early detection of outbreaks, but timely availability of codes is a problem in civilian systems. The outpatient Department of Defense ESSENCE system is an example of a large-scale surveillance system based on ICD-9 coded diagnoses. ESSENCE had detected outbreaks at military treatment facilities, primarily respiratory and gastrointestinal. Many of these outbreaks have been validated with investigations by preventive medicine staff and occasionally, with clinical culture results (example San Diego GI, 2002, Pavlin, personal communication).

3.5 *Chest Radiographs*

In the United States, radiologists dictate or otherwise record a description and interpretation for most chest radiograph and these dictations are transcribed into electronic format. Some hospitals have even incorporated speech recognition systems in the radiology department to create electronic reports at the time of interpretation, however, reports at most hospitals are transcribed after dictation and are available electronically with a 12- to 24-hour latency

3.5.1 Data Characteristics

The transcribed reports describe—in free text format—specific findings important for detection of infectious diseases of the lower respiratory tract such as SARS, Plague, Tularemia, and inhalational Anthrax. The granularity of the information is quite specific and allows for detection of different patterns of pneumonia, pleural effusions, and mediastinal widening. The data are identified at the level of the individual patient and can therefore be linked to home location, or they can be linked with other data about the patient to achieve more specific case definitions, and analyzed spatially with data from other patients to detect clusters of cases, or the data can be processed after de-identification to protect privacy. Previous NLP work has focused on encoding findings from chest radiograph reports and has identified findings such as atelectasis, localized infiltrates, consolidation, pneumonia, and pneumothoraces as accurately as physicians [35-38]

3.5.2 Experiments

Applications for detecting patients with radiological evidence of mediastinal findings consistent with inhalational anthrax and of acute bacterial pneumonia have been developed and evaluated. As part of Bio-ALIRT, Chapman and colleagues used the IPS machine learning system to create a keyword-based detector that detected patients with mediastinal evidence of anthrax with a sensitivity of 0.35 and a specificity of 0.99 [39]. A modified version of the system increased sensitivity to 0.86, maintaining a sensitivity of 0.99. A keyword search that accounts for negation detected patients with evidence of pneumonia with a sensitivity of 0.85 and a specificity of 0.96.

The latency at which chest radiograph reports (in 2001) were available electronically on the MARS hospital information system at the University of Pittsburgh Medical Center averaged of 16 hours after the radiograph was performed.

3.5.3 Conclusions about Timeliness and Detectability

Radiology reports are not as timely as chief complaints; however, the information obtained is more specific. Detecting a small outbreak of patients with respiratory diseases such as SARS and inhalational anthrax may require specific information such as that which can be found in the chest radiograph report. Natural language processing techniques can accurately encode findings from chest x-ray reports. The encoded findings could then be made available to syndromic surveillance systems to allow them to count cases of, for example, fever and pneumonia.

3.6 From Permissive Environments

Certain types of "permissive" environments may allow collection of data that are not easily available in more general contexts. Examples of permissive environments include some types of worksites, university campuses, or military bases. They are typically characterized by a geographically constrained location (a "site"), some centralized control of IT and telecommunications infrastructure, as well as a shared sense of community and trust. The centralized control of IT/telecommunications allows electronic tracking of data (e.g., outgoing phone calls) that are difficult to obtain on, say, a city-wide basis. The shared sense of community and trust makes it easier to gather personal health information, as there is a belief that these data will not be used to violate privacy rights.

3.6.1 Data Characteristics

Multiple types of data are available in permissive environments, so data characteristics are described in the experiments.

3.6.2 Experiments

3.6.2.1 Polling Surveillance (Watson Health Survey)

As part of Bio-ALIRT, The Watson Health Survey was run at the IBM T.J. Watson Research Center from 1/25/2002 through 5/31/2002 [40]. The Watson Research Center is made up of two sites: Yorktown Heights, NY and Hawthorne, NY. The sites are approximately 10 miles apart. Watson employees were asked via email to participate in the survey, and 397 employees (out of approximately 2200) agreed to participate. Each participant was sent email once a week at 7 a.m. on a randomly chosen work day. Participants were asked to self-assess their health for the day the email was sent, choosing from one of the following categories: As usual, Worse than usual (sick), or Much worse than usual (very sick). If the participant reported being sick or very sick, there was a series of additional questions about the illness, including a question about symptoms (fever, headache, muscle ache, fatigue, cough, stuffy/runny nose, etc.). During the course of the survey there were minimal dropouts (< 2%), and responses were very timely: 92% of the responses showed up the same day, 73% by noon. After a survey response was received and added to the aggregate totals, the response was deleted to ensure privacy.

The survey data were evaluated by examining the correlation with seasonal disease activity (see Box 3.1 for background on correlation analysis). The gold standard data used were physician's office visits with respiratory-related ICD9 codes (460-519) in the Westchester County area. The office visits were determined from insurance claims data and the coverage for the area was estimated to be 20%. We compared office visits with the percentage of respondents that reported being sick. The correlation between these two data sources was high, with the maximum correlation of .82 occurring with survey leading office visits by three days [17, 40].

3.6.2.2 Telephone calls to medical offices

The Watson Research Center tracks outgoing phone calls for billing purposes. As part of Bio-ALIRT, a list of approximately 5000 phone numbers was compiled that corresponded to medical offices in the area surrounding the Watson sites (Westchester County). These numbers were obtained from various sources including both online and hardcopy phone directories. A daily report was obtained of the total number of outgoing calls from each site to medical offices, as well as the number of callers (extensions) making those calls. These totals have been recorded beginning in October 2001.

The telephone call data were evaluated by examining the correlation with seasonal disease activity. The gold standard data used were physician's office visits with respiratory-related ICD9 codes (460-519) in the Westchester County area. The office visits were determined from insurance claims data and the coverage for the area was estimated to be 20%. We compared office visits with total number of callers to medical offices for the time period October 2001 through May 2002. Weekends were ignored, as call volumes drop dramatically on non-working days. The maximum correlation between these two data sources was .72, which occurred when telephone calls led office visits by four days [17, 40].

3.6.3 Conclusions about Timeliness and Detectability

There is preliminary evidence that suggests that self-assessed health surveys and telephone calls to medical offices lead physician office visits by three and four days respectively. These results need to be confirmed in other locations, over longer periods, and using alternative evaluation methodologies.

4 OTHER DATA SOURCES

The previous section described non traditional data with the most potential for early detection of disease outbreaks-- ICD-9 diagnoses, free-text chief complaints, chest radiographs, self-reporting, phone call traffic, and OTC sales. Disease outbreaks may also leave early footprints in other types of data and a key goal of BioALIRT was to obtain and explore as many types of data as possible. This section describes those additional investigations

4.1 Orders for Laboratory Tests

Orders for laboratory tests may provide an electronically available clue to the diagnostic thinking of clinicians and importantly precede the results of those tests by several days. Test requests may suggest a general category of illness, as influenza tests are associated with respiratory illness, or they may be much less specific, as complete blood cell counts may be ordered in connection with nearly any condition. But because clinicians may not order tests for every condition they encounter or on every occasion, test request data is less comprehensive than visit associated data such as ICD-9-CM codes or chief complaints.

4.1.1 Data Characteristics

Orders for laboratory tests from emergency departments and inpatient facilities are commonly electronic; laboratory tests ordered from outpatient facilities are less commonly so. Like radiology results described in Section 3, the data are identified at the level of the individual patient and can therefore be linked to home location and analyzed spatially with data from other patients to detect clusters of cases, or they can be linked with other data about the patient to achieve more specific case definitions, or the data can be processed after de-identification to protect privacy. The names of laboratory tests have been standardized by the LOINC coding system which is not ubiquitously in use but is used by national laboratory companies and an increasing number of vendors of laboratory information systems. For those entities that do not utilize LOINC encoding, expensive translation tables must be developed to map from the proprietary codes to LOINC codes for the data to be usable in regional biosurveillance.

4.1.2 Experiments

As part of Bio-ALIRT, state health department laboratory influenza test results and orders (Figures 4.1 and 4.2b) were analyzed by Chotani and Lewis of JHUAPL and Coberly of JHUSPH who found high correlation between influenza test results and the onset of the influenza outbreak during the past four seasons (personal communication Joe Lombardo). There was, however, no lead-time for test requests compared to patterns of influenza-like illnesses ICD-9 diagnoses in physician office visit record (Figure 4.2). This result suggests that either source can be used for monitoring and the preference between them depends on whether the data can be obtained without time delay and differences in cost and effort to build the data collection system. A caveat to this research is that the data source investigated was limited to tests ordered from the state laboratory whereas influenza testing is also done by private laboratories, which may produce an earlier or stronger signal (due to increased sampling).

There are other factors to be considered such as the age distribution for tests. In the ESSENCE data, the majority of samples were collected from elderly patients, probably because the elderly are more at risk. But are the elderly good sentinels for bio terrorism? A study of nursing home data collected for ESSENCE during the 1999-2000 influenza season showed that residents were not the first cases to show up in that setting, but followed the staff illnesses by one to two incubation periods (Joe Lombardo, personal communication).

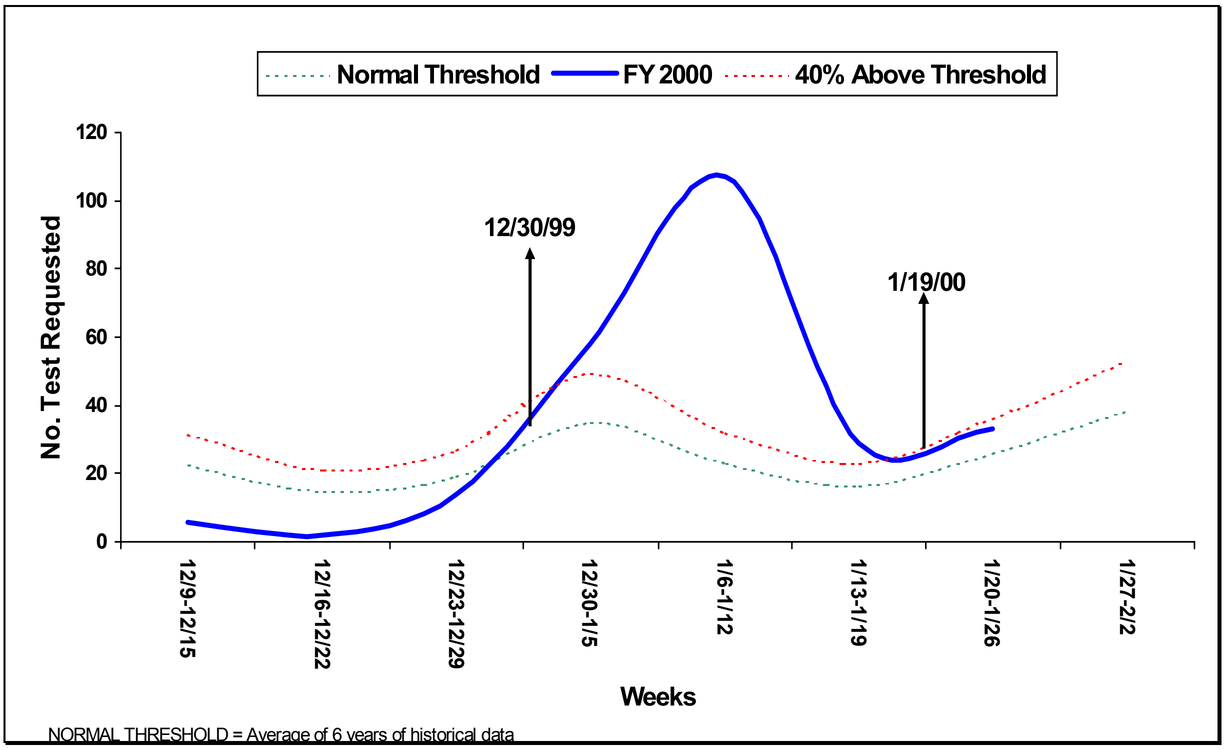

**Figure 4.1: Influenza Lab Test Requests:
Weekly Distribution of Influenza Tests
Requested by Specimen Collection Date
12/9/00 to 1/26/00**

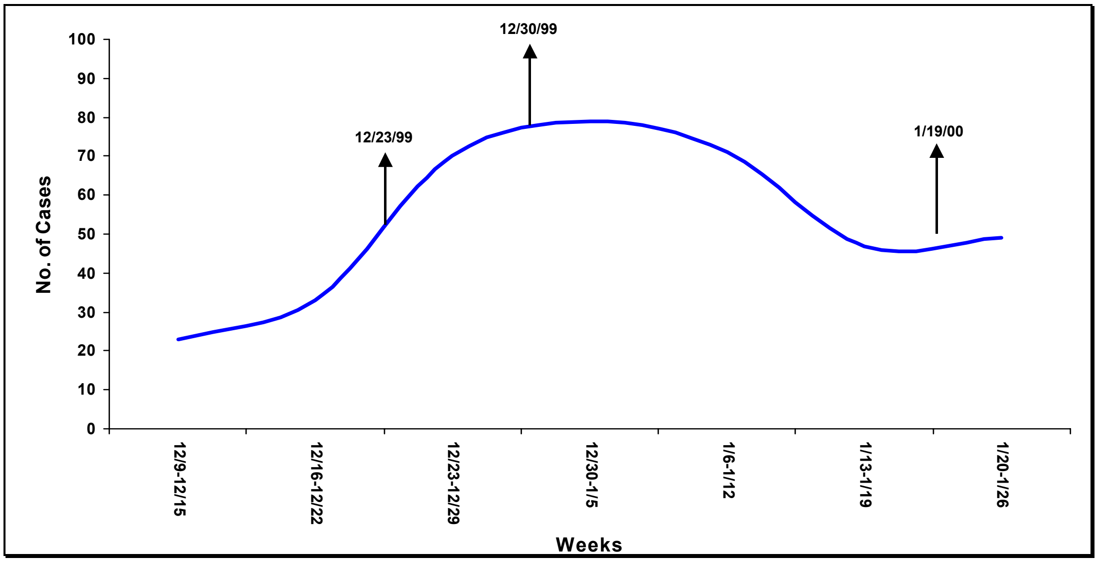

Figure 4.2: Emergency Room Chief Complaints for ILI:
Weekly Number of ER Cases with ILI Influenza or Pneumonia
County Hospital 12/9/99 to 1/26/00

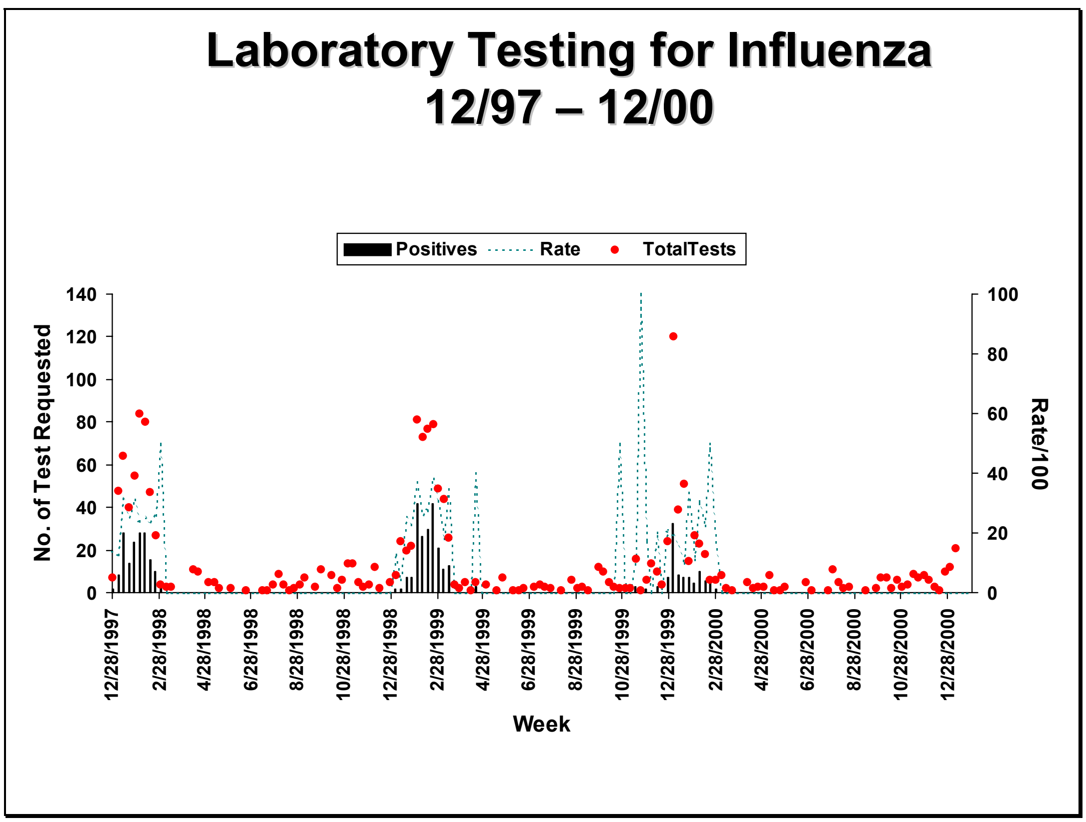

Figure 4.2b: Time distributions of influenza test lab requests and influenza test positives, by specimen collection data. "Rate" refers to the ratio of positive results to requests.

4.1.3 Conclusions
Influenza test requests are an available data source and correlate with the onset of influenza season. Nevertheless, this study did not demonstrate that they provide a timely signal as a leading indicator. Since test requests are made at the time of a patient encounter they may not be more timely than visit-related data such as ICD codes or chief complaint information if the clinical practices monitored submit billing claims on the day of service or immediately afterward. Finally, testing of medical staff, as shown in the nursing home data, may provide a timelier signal than testing of patients.

This data was limited to tests ordered from the state laboratory. Most influenza testing is done by private, rather than the state laboratories. Analysis of private laboratory data might provide different results.

4.2 Poison Centers
Dissemination of information related to poisons is the primary role of poison information centers. The information is provided by specialists in poison information (nurses and pharmacists) and clinical toxicologists in response to calls from the public or medical facilities. Each interaction is documented electronically and constitutes the poison center medical record. Poison information centers collect a large volume of information on exposures involving both humans and animals. For example, the Pittsburgh Poison Center responded to approximately 78,000 inquiries in 2000. Nearly 5,000 of the inquiries involved animal exposures.

4.2.1 Data Characteristics
The poison center medical record contains personal and demographic information about the patient as well as specific information that are categorized by substance (the poison), treatment, patient symptoms, and route of exposure and laboratory values. The values within each section are fully searchable and standardized by national convention so that data from all poison centers in the United States can be incorporated into a single database. There is also a free-text documentation section. Currently, data are submitted to the American Association of Poison Control Centers Toxic Exposure Surveillance System (AAPCC TESS) on a monthly, quarterly or semi-annual basis by participating poison centers. Participation in AAPCC TESS and submission of all data is mandatory for Certified Regional Poison Information Centers and voluntary for non-certified centers. Data are reviewed and published on an annual basis approximately nine months after the end of the calendar year. Sixty-three poison centers submitted data to AAPCC TESS in 2000. Only Alaska, Hawaii, Mississippi, northwest Ohio, South Carolina and Vermont are not represented by the AAPCC TESS data.

4.2.2 Experiments
The data were inspected for evident spikes. Figure 4.3 shows a prominent increase in the proportion of calls on Thanksgiving, 2001, relative to the average for calls for diarrhea, nausea and vomiting. This spike, however, is the only biological event noted in poison center data in several years suggesting that sick individuals do not call poison centers unless they suspect that their sickness is due to an accidental ingestion.

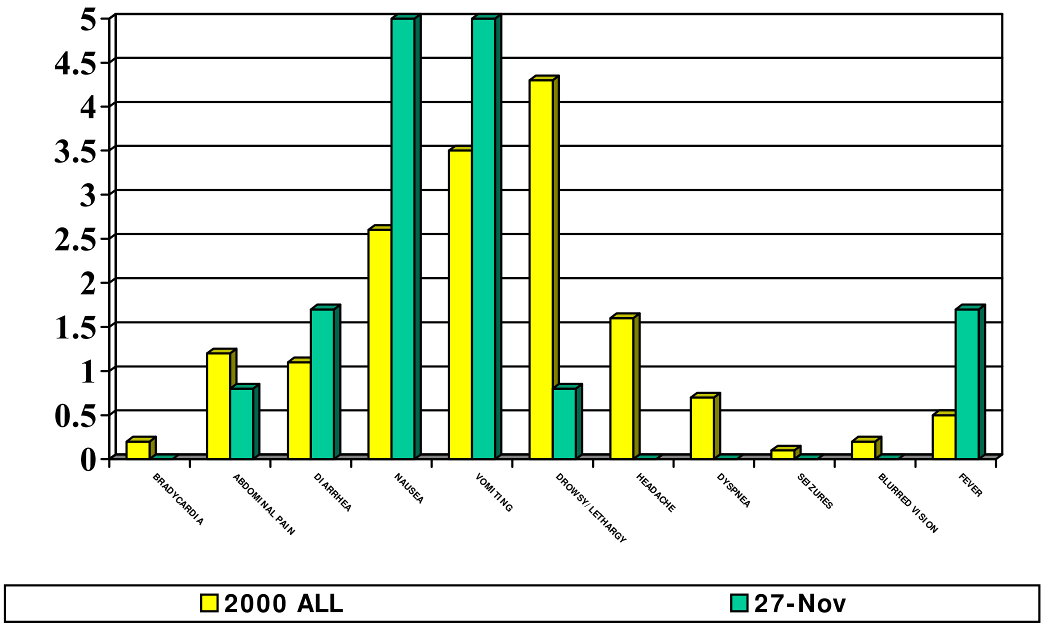

Figure 4.3 Daily Surveillance report from Pittsburgh Poison Center for November 27, 2001, percent of calls normalized by total 2000 calls

4.2.3 Conclusions

Because people may not contact poison centers when they are ill unless they suspect a specific substance as the cause (or they wonder whether a child may have ingested something) the ability to detect infectious disease outbreaks is likely limited.

4.3 Waste Water

Chemical measurements of waste water constituents are made routinely in major cities to control the waste water treatment process. These data are recorded routinely and available for analysis.

4.3.1 Data Characteristics

They are typically recorded on a regular basis and can be available for analysis. They are not typically machine readable.

4.3.2 Experiments

To evaluate wastewater data as an early population outbreak indicator, Brinkman conducted a retrospective analysis of Milwaukee wastewater chemistry data collected in during a cryptosporidium outbreak in 1993. The outbreak occurred during a period from about March 18 to April 27, a period that was associated with very high rainfall and affected 400,000 individuals [9-11]. The precipitation and associated runoff from agricultural areas may, in fact, have caused the outbreak. The public health department became aware of the outbreak on April 5, 2003.

Routine data collected at two treatment plants at the time of the outbreak were available for analysis. The data are limited to daily flow and daily composite samples for eight constituents.

From data collected daily from October 1, 1992 through September 30, 1994 we calculated the mass of the measured constituents by multiplying the daily concentration of each constituent by the daily inflow volume. We normalized the results to zero mean and unit standard deviation and removed the first order dependency on dilution from precipitation.

Confounding factors, including changes in manmade use patterns, precipitation, and wastewater diversions, were assessed. Sewage system use patterns did not change significantly around the time of the outbreak and their effects on wastewater flow and chemistry were inconsequential. Varying diversions between the two treatment plants serving the city occurred as a result of rainfall. Therefore, evaluating the effects of rainfall patterns indirectly assesses the affects of these diversions. As shown in Figure X and in detail in Figure X+1, changes in rainfall were not associated with changes in the daily phosphorus masses. Ammonia masses tended to decrease when the precipitation increased.

Among the measured constituents, the mass of phosphorus and of ammonia correlated with the outbreak. In particular, the mass of phosphorus rose to over three standard deviations above the normalized mean for a period of five sequential days during the outbreak, an event not found elsewhere in the data reviewed. Figures 4.4 and 4.5 show the remarkable excursions of the normalized weekly sum of ammonia and phosphorus masses handled by the two sewage treatment systems service Milwaukee during the outbreak (marked).
:

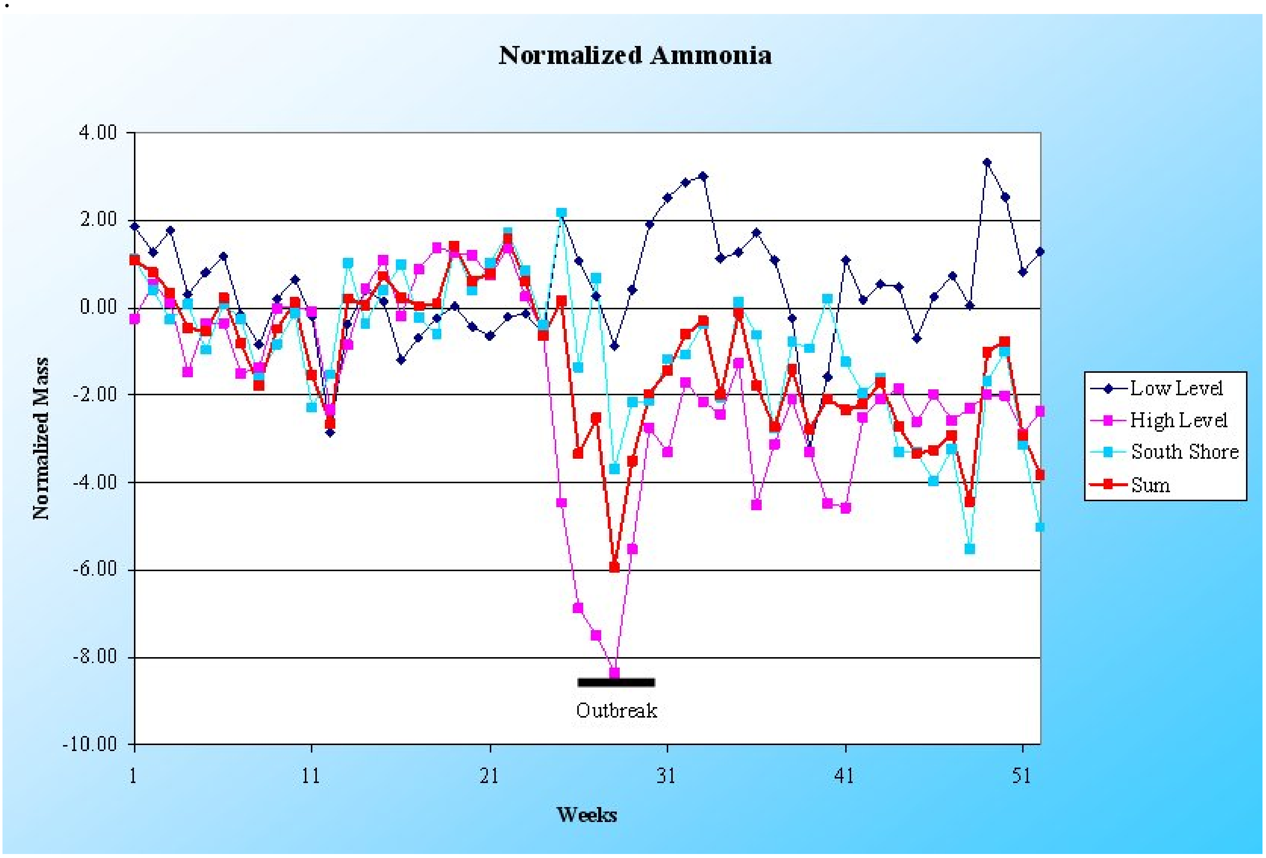

Figure 4.4. Decreased Sum of Ammonia Mass during the Cryptosporidium Outbreak

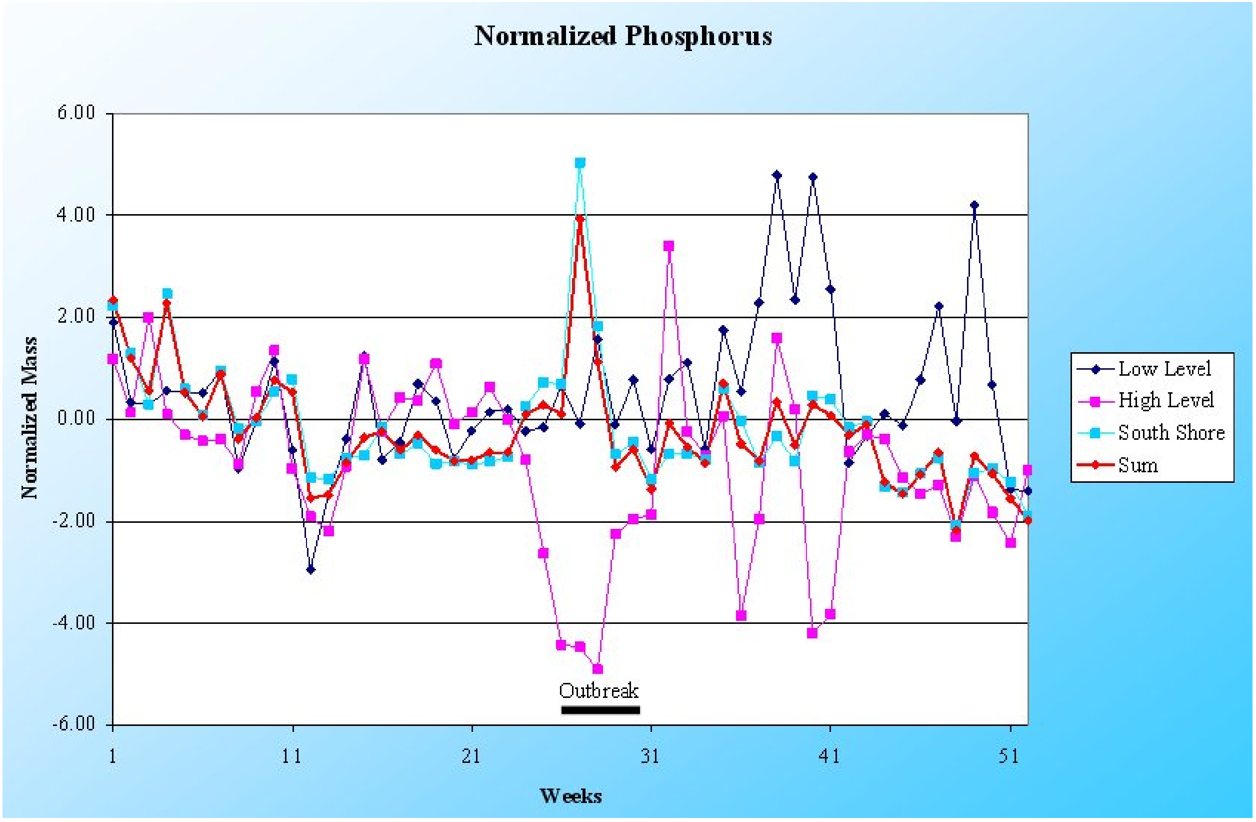

Figure 4.5 Increased Sum of Phosphorus Mass during Cryptosporidium Outbreak

4.3.3 Conclusions

A significant increase in the phosphorus mass and decrease in ammonia mass carried in wastewater occurred during the 1993 cryptosporidium outbreak in Milwaukee. The phosphorus peak occurred nearly three weeks after the outbreak began, and about a week after the public health department became aware of the community illness, suggesting that existing methods of detection are superior for an outbreak of this organism. Further work with other similar outbreaks is necessary to confirm the reliability relationship and its timing.

4.4 School Absenteeism

School absenteeism may be an early indictor of illness in school-aged populations.

4.4.1 Data Characteristics

School absentee rosters can be obtained electronically in some jurisdictions. There is variation, however, in how these data are collected and stored. In the ESSENCE National Capital Area system, school absentee rosters can often be made obtained electronically by the next day however this degree of timeliness may not be achievable in other regions. An attractive attribute of this data source, when available, is that it samples almost the entire school-aged population.

A disadvantage is that the data are non-specific; the reason for absence is generally not available and there are blackout periods corresponding to weekends, holidays, and summer recesses.

4.4.2 Experiments

Daily absence numbers were collected by Vasholz and colleagues [Vasholz et al, see Appendix] from over 100 elementary schools in 2 Maryland Counties daily over a 4-year period, together with monthly enrollment totals. They measured correlations between elementary school absenteeism and physician office diagnoses obtained from insurance claims and also between absenteeism and hospital emergency department discharge diagnoses. These correlations are measured after removing predictable outliers (e.g. near holidays like Valentine's Day and near spring vacation, etc.) in the absentee data. The results show correlations in the range of 50% to 80% between clinical visits for acute illnesses and elementary school absentees (Figure 4.6). Anecdotally, it has been observed that data from middle and high school students do not correlate as well.

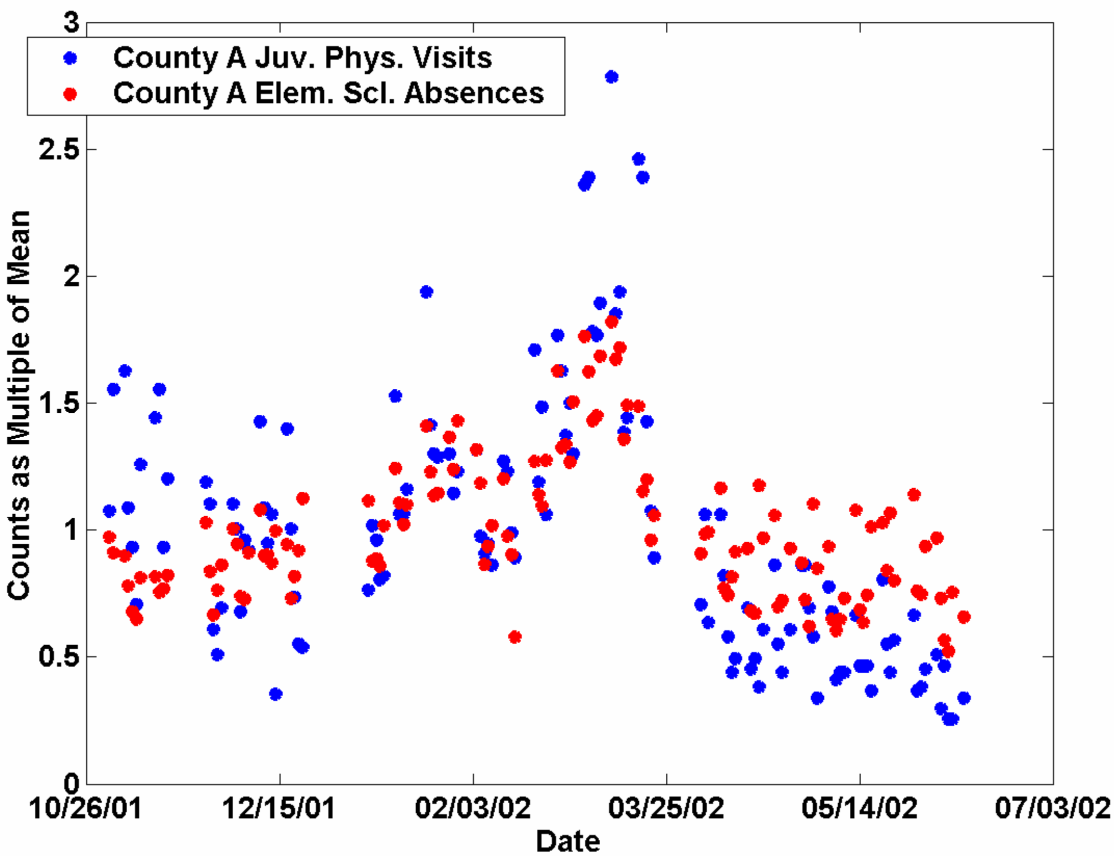

Figure 4.6. Relationship between Elementary School Absences and Physician Visits for Juveniles (counts as multiples of the mean)

4.4.3 Conclusions

School absentee data, when available electronically, appears to represent a timely source of information regarding the health status of children, especially elementary school-aged children. The sampling rate of the school-aged population is obviously very high when all schools participate, but the information derived from absentee data alone (i. e. no reasons obtained for the absences) is non-specific regarding the nature of any health problems that may be observed.

4.5 Animals

Animals usually get different diseases than humans because pathogens adapt to specific host species. Important exceptions are zoonotic diseases, those that infect both animals and humans. Most of the diseases that are bio-terrorist threats are also zoonotic. Zoonotic diseases occur naturally and most commonly in regions with poor public health resources and infrastructure. They are relatively rare in the US but they are important when they do occur (e.g., rabies, West Nile virus). Active monitoring the health of animals to provide warning of human health-threatening pathogens depends upon specimen collection and serological testing or necropsy, which may take several days or longer. Therefore, passive monitoring could provide earlier indications of a disease outbreak. In particular, knowledge of the typical animal disease background by passive monitoring should assist in the detection of an anomalous spike in animal diseases.

Animal species are potentially useful as timely sentinels in syndromic surveillance if they are exposed to the disease simultaneously with humans, exhibit obvious disease signs, and have greater susceptibility and shorter incubation periods than humans for the disease. While these requirements limit the usefulness of animal monitoring, there remain a number of animal species that do meet these requirements, thereby providing potentially useful passive sentinels. For example, sheep, cattle, goats and horses are more susceptible to anthrax than humans [41, 42]. Cats, rats, prairie dogs, chipmunks, and ground squirrels may be useful as sentinel animals for plague [43, 44]. The 1994 pneumonic plague outbreak in India was preceded by a large urban rat die-off a few weeks earlier than the first reported human case (CDC 1994, John 1994). Rohrbach suggested that local die-offs of rabbits or rodents should be investigated for possible tularemia [45].

4.5.1 Data Characteristics

The minimum data characteristics for all types of animal sources include number of animals, species, date of examination, location, and syndrome. The specific data characteristics will vary by source. The location may be the zip code or county name of the pet owner, farm, ranch, care facility, place of examination, etc. The syndrome may be entered directly by the animal health professional or derived indirectly based on the primary presenting signs of the animal. Because many veterinary health professionals still maintain primarily paper health records and any animal health databases are either small, specialized or non-existent, the establishment of health databases similar to that found for humans has proved to be a challenge. Furthermore, such data must be collected consistently over a long period of time in order to ascertain normal periodic patterns.

4.5.2 Experiments

As part of Bio-ALIRT, Babin and colleagues monitored syndromic information from veterinary clinics, animal control facilities, and wildlife rehabilitation centers. They found it possible to track syndromic information from animals and even detected a large anomaly that resulted from a weather-related cause. However, because the establishment of historical animal health databases is in a very early stage, evaluation of the utility of passive monitoring in predicting human outbreaks cannot yet be determined (Babin, S., J. Casper, C. Witt, S. Happel Lewis, R. Wojcik, S. Magruder, H. Burkom, J. Weitzel, and J. Lombardo, Early detection of possible bioterrorist events using sentinel animals. Abstract available, presented at the 131st Annual Meeting of the American Public Health Association, San Francisco, CA, 15-19 November 2003).

4.5.3 Conclusions

Early detection of bioterrorism by passive surveillance of animals poses unique challenges. The pathogen may be introduced directly into urban areas where many potential animal sentinels are likely to be absent. The large historical databases of health data that exist for humans because of insurance requirements are not typically kept for animals. For all these reasons, the role of animal surveillance to provide early warning of zoonotic diseases that are intentionally introduced into the human population is uncertain at this time.

4.6 Prescription Drugs

Tracking prescriptions could have utility as a tool in real-time surveillance. The information is regularly captured and available. Few researchers have evaluated prescription drugs for utility as a tool in real-time surveillance.

4.6.1 Data Characteristics

The Pharmacy Data Transaction Service (PDTS) monitors all prescriptions paid for or reimbursed by Tricare, the military health system. The information is collected electronically at the time the prescription is filled, and arrives in the electronic database with 3.2 seconds. Many drugs that can be purchased over-the-counter in civilian settings are prescribed at military treatment facilities (MTFs), such as decongestants and anti-diarrheals, and they are also included in the data.

4.6.2 Experiments

As part of BioALIRT, in the spring of 2002, the DoD ESSENCE program was able to evaluate the usefulness of military prescriptions. Using the National Capital area as a test bed, prescriptions written on the same day as an ambulatory visit were abstracted from two linked databases: the Ambulatory Data Module (ADM) and the PDTS. Applicable drugs prescribed at those visits were placed into either a respiratory or GI group based on the decision of a preventive medicine physician. Drugs were further grouped into broad GCN groups based on an existing standard pharmacy scheme. Prescription counts by syndrome group were graphed over time and correlated with like ADS ICD9 syndrome groups from the ESSENCE system.

Six of the GCNs groups were highly correlated with ADS respiratory group and 2 GCNs correlated well with ADS GI (Figure 4.7) (Eader, unpublished data).

A separate study found that antiviral drugs and ADM respiratory visits were highly correlated but there was little lead-time gained with PDTS. (Foster VB and Elbert E, unpublished data). In addition, PDTS antianxiety and antidepressant prescriptions were highly correlated with mental health visits for depression and anxiety [46].

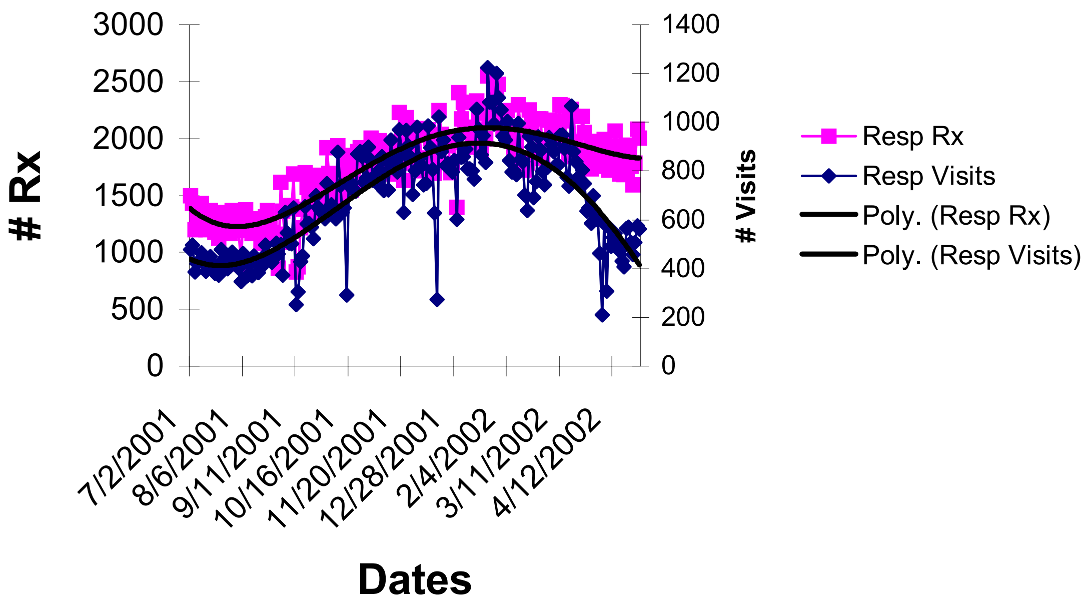

Figure 4.7. Example of Respiratory ICD9 Syndrome and Respiratory GCN Group Correlation

4.6.3 Conclusions

Prescription data correlates well with clinical visits. Depending on reporting procedures at the sources, increases in prescription fills may occur before increases in outpatient ICD9 code reports and therefore allow for earlier alerting of increased disease in a population. Additionally, based on results from mental health studies done at Walter Reed, prescription data has been shown to accurately track non-infectious events and can potentially provide timelier event detection. While preliminary studies have shown prescription data to be an early indicator, more analyses need to be accomplished. Further research will be done at Walter Reed to refine drug groups by comparing prescription data to known military outbreaks.

4.7 Web Queries

Surveys of Internet utilization suggest the potential of monitoring health-related websites for outbreak detection. Recent reports estimate that from 16.1% to 40% of Americans already look for health information online [47, 48]. Surveys by the Pew Internet & American Life Project report that the "Internet population" – defined as those Americans with online access who identify themselves as computer users – has stabilized at about 60% of Americans since 2001, but the number of experienced users has grown significantly suggesting that with time, the proportion of the sick seeking information online may increase [47].

4.7.1 Data Characteristics

Web access logs are maintained electronically by most Web sites and contain a variety of information about users and information users accessed. Data available may include:

- IP address – the IP address of the computer making the HTTP request
- RFC – a field that is used to identify the person requesting information from the Web site. (This field was null in our Web access logs.)
- Auth – a field included to list the authenticated user, if required for Web site access
- Timestamp – providing date and time information to the second
- Action – the action requested by the user (e.g., request for a specific article or for a search)

Status – a code indicating whether the user experienced a success, redirect, failure, or server error when they tried to access the Web page.

- Transfer Volume – indicating how many bytes were transferred to the user.
- Referring URL – the URL of any Web site that the user was using just before coming to the present site.

4.7.2 Experiments

As part of BioALIRT, Johnson and colleagues studied 12 Web access logs (one for each month of the 2001 calendar year) obtained from Healthlink, a consumer health information Web site, developed and maintained by the Office of Clinical Informatics at the Medical College of Wisconsin as a service to their patients and community, to determine whether they contained a signal of influenza activity and whether any signal found was more timely than conventional influenza surveillance data [49]. They used the cross-correlation function (CCF) to measure correlation between weekly counts of accesses to 17 "influenza-related" articles on the Healthlink site, and two gold standards: the number of influenza-like-illness (ILI) cases reported by U.S. Influenza Sentinel Physicians Surveillance Network and the number of positive influenza cultures reported to the Centers for Disease Control and Prevention (CDC). Timeliness was defined as the time lag at which the correlation was a maximum. The analysis was limited to those time periods where the Healthlink data and influenza surveillance data overlapped: namely, portions of the 2000-2001 and 2001-2002 influenza seasons, weeks 1 – 20 and 40 – 52, respectively. The correlation between the Article Total time series and the ILI gold standard was 0.78 for weeks 1 – 20 of 2001 and 0.76 for weeks 40 – 52 (Figure 4.8).

The performed cross-correlation analysis (see Box 3.1) of a group of 17 "influenza-related" articles with the following influenza surveillance data from the Centers for Disease Control and Prevention as gold-standard: the number of influenza-like-illness (ILI) cases reported by U.S. Influenza Sentinel Physicians Surveillance Network and the number of positive influenza cultures reported to the CDC. Timeliness was defined as the time lag at which the correlation was a maximum. They performed the analysis for those time periods where the Healthlink data and influenza surveillance data overlapped: namely, portions of the 2000-2001 and 2001-2002 influenza seasons, weeks 1 – 20 and 40 – 52, respectively.

The correlation between the Article Total time series and the ILI gold standard was 0.78 for weeks 1 – 20 of 2001 and 0.76 for weeks 40 – 52 (Figure 4.6). Timeliness of this time series was zero for both time periods, meaning neither time series (Article Total nor ILI gold standard) preceded the other. When compared to the positive cultures gold standard the correlation was 0.67 and 0.80. Although the Article Total time series lagged three weeks behind the positive cultures gold standard during weeks 1-20, it preceded that gold standard by two weeks during weeks 40-52.

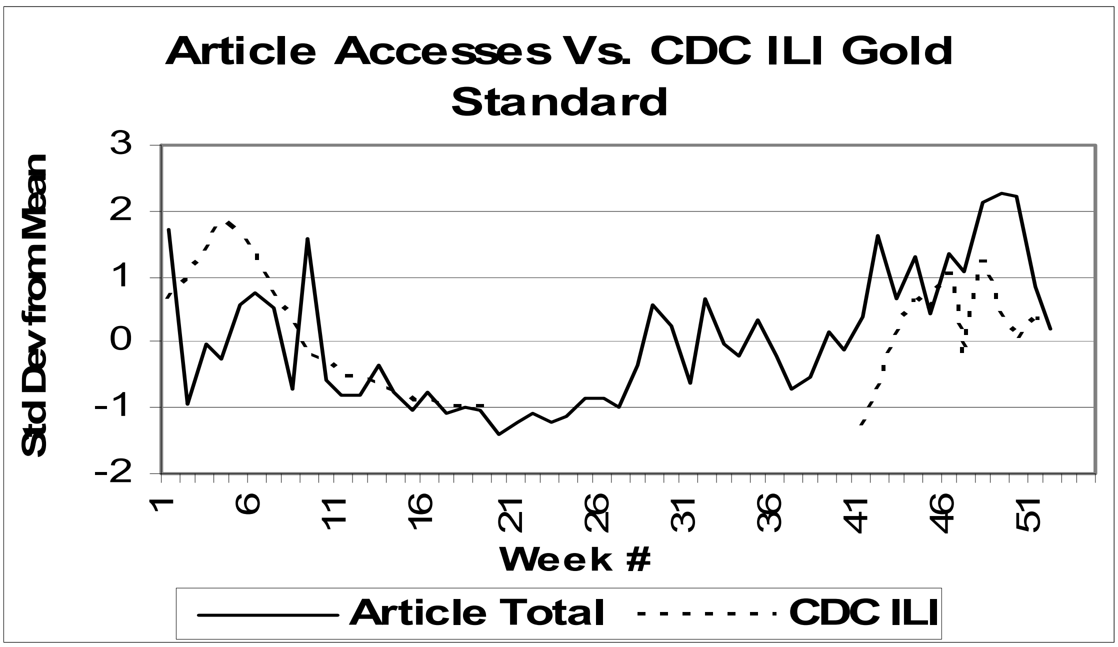

Figure 4.8 Correlation of Influenza article accesses on the Web and Influenza activity as measured by standard surveillance datasets. 2001.

4.7.3 Conclusions

User access to articles related to influenza, as documented in Web access logs, appears to be a signal for influenza outbreaks. Timeliness results were variable and illustrate that consumer searches for medical information may precede or follow medical encounters. Extending this study to include additional time periods, looking at information from other health related web sites (e.g. PubMed), using symptom information extracted from free text queries at health related web sites, and employing sophisticated methods for performing spatial analysis may further demonstrate the potential of this source as a useful public health surveillance tool.

4.8 Weather and other Environmental

Environmental factors such as weather and pollution impact human health and influence bio-surveillance data such as patient visit counts. For example, the peak occurrence of respiratory illness occurs seasonally during the coldest day of the year [50]; the Centers for Disease Control and Prevention documented significant increases in diarrhea and stomach ailments in areas flooded by Tropical Storm Allison [51]; air pollution is associated with increases in hospital and emergency department admissions [52, 53] and in increased sales of ambulatory respiratory drugs [54]. Accounting for the contribution of such factors to naturally occurring illness is important to the interpretation of bio-surveillance data and the determination of alert levels.

4.8.1 Data Characteristics

Information such as daily maximum, minimum, and mean temperatures; and amounts of rainfall and snowfall at airports is regularly recorded by the National Weather Service and available on public web sites. Information about pollen counts and air pollution indices is similarly available

4.8.2 Experiments

Comparison of National Weather Service data with clinical data showed that peaks in human respiratory illness tended to follow cold air mass movements across the United States Babin [50]. Florio and others showed that outdoor temperature was a better surrogate than calendar date for explaining seasonal variations in sales of over-the-counter (OTC) medications [16,22] and that outdoor temperature proved a valuable predictor for anti-flu OTC sales in the National Capital Region.

4.8.3 Conclusions

Environmental data can help explain observed variations in human health data obtained in support of bio-surveillance. Accounting for the influence of weather variables such as temperature will be valuable tools in improving the reliability of alerts to possible bio-terrorist activities based observations of other data streams

4.9 Mass Transit Utilization

Illness in major metropolitan areas may be reflected as a drop in ridership of commuter trains, subways, buses, or other public transit. Shah and colleagues evaluated whether Pittsburgh bus ridership (PBR) data can potentially signal an outbreak.

4.9.1 Data Characteristics

PBR data consist of date, route identification, route description, mode of payment (e.g., cash, weekly or monthly pass holders) and type of rider (e.g., senior, student, and handicapped) in Allegheny County from Nov 2002 to Oct 2003. The ridership information is stored in the fare box for each day and automatically downloaded to a database in a standard format early next morning. Data quality check programs are run before importing the data into the database. Moreover, PBR captures different age segments of the population and according to the 2000 US Census, an estimated 11-16% of Allegheny county residents above the age of 16 years use public transportation.

4.9.2 Experiments

Shah et al. (data cannot be published due to restriction by data provider) converted daily counts of PBR to weekly counts and compared them to a gold standard of the percentage of positive influenza tests per week in the Mid-Atlantic region from Nov 2002 to May 2003 as posted on the CDC website to evaluate the timeliness of PBR as an early detector of influenza outbreaks. Time latency was measured by the cross-correlation function (CCF). Separate time series were created for weekly or monthly pass holders and seniors.

CCF showed a four week lead of PBR over the influenza test gold standard. There was moderate correlation between senior ridership and the gold standard, and poor correlation between weekly or monthly pass holders and the gold standard (see Figure 4.9). One could speculate that weekly and monthly pass holders, often people commuting to employment, were less likely to elect not to travel than were seniors who were not commuting to work.

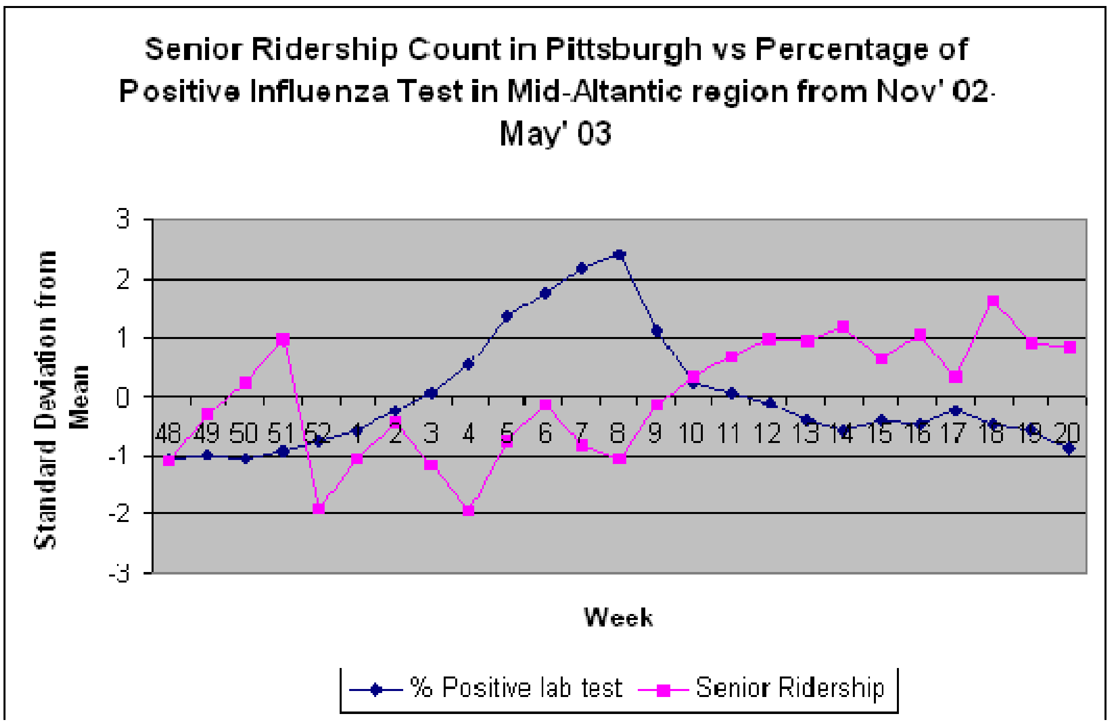

Figure 4.9. Weekly total bus ridership by seniors and Influenza activity. Pittsburgh, Pennsylvania.

4.9.3 Conclusions

Senior PBR appears to be a promising marker for early detection of influenza outbreaks. The small number of data points and the course granularity of the gold standard limited tests for statistical significance. Further studies may indicate that the behavior of seniors as reflected in public transit ridership, are important sentinels for some categories of illness. Studies of route information could be helpful in detecting localized outbreaks.

Furthermore, future studies need to explore possible confounding factors such as weather and road construction and examine the value of route information for detecting localized outbreaks.

Bus ridership may not be a good representative of other types of mass transit utilization data. A key limitation of bus ridership is that the counts are collected at the level of route, not bus stop. Subway and Metro systems potentially can provide entrance and exit data for each stop, supporting a more fine-grained geographical analysis and potentially the ability to detect outbreaks confined to large buildings or neighborhoods served by specific metro stops.

4.10 Cafeteria Sales

People suffering symptoms of various illnesses may alter their patterns of purchase of foods in locations where some observation is possible, such as at cafeterias. For example, respiratory illness may cause an increase in sales of soups or beverages, both foods commonly considered helpful or soothing in such cases. Such a change in pattern could be a timely indicator that the population suffers symptoms not yet reported to clinicians.

4.10.1 Data Characteristics

Itemized cafeteria sales data are often collected electronically from point of sale (POS) devices Data collection is immediate and stored in machine-readable databases. These data yields quantities of food items purchased, the number of purchasers, and, typically, price. Such data could be segmented over time periods throughout the day.

4.10.2 Experiments

As part of BioALIRT, IBM examined the purchases of beverages during the first three months of 2002 in the cafeterias at the IBM T.J. Watson Research Center, located in Yorktown Heights, NY and Hawthorne, NY [17].

Campbell and colleagues examined the purchases of beverages during the first three months of 2002 in the cafeterias at the IBM T.J. Watson Research Center, located in Yorktown Heights, NY and Hawthorne, NY. [17]. They evaluated the correlation of total beverage sales data with a gold standard data of physician's office visits with respiratory-related ICD9 codes (460-519) in the Westchester County area obtained from insurance claims data covering the period January-March, 2002. Weekends were ignored because the cafeterias were closed on these days, and Mondays were ignored because beverages were free and sales not recorded. The maximum correlation was 0.66, with a lead-time 19 days.

4.10.3 Conclusions

Although the investigators considered the preliminary study of cafeteria data to be inconclusive (due to the biologically implausible 19-day lead time), there appears to be sufficient evidence to warrant further investigation. Many possible confounding factors in addition to those considered may be important. Normalization of sales for the number of patrons could be useful, and increased use of beverages by symptomatic patients may be particularly reflected in sales outside usual meal hours and, perhaps, outside the cafeteria. Other food items such as fresh fruit or soup at worksite cafeterias may demonstrate more helpful correlations.

4.11 Parking Lot Utilization

Overall increased use of medical facilities may be an early indicator of increased illness in a population. Nonspecific indicators such as use of parking facilities connected with medical facilities use may provide more timely information than clinical visit data from outpatient clinics or emergency rooms.

4.11.1 Data Characteristics

In connection with collecting fees, some parking facilities associated with medical care facilities control access and automatically collect usage information. The information may distinguish categories of users (cash payers, pass users, third party payers, etc). The data may be collected electronically and available on a daily basis. Direct data collection from electronic data bases is possible.

4.11.2 Experiments

Stetson and others (unpublished data presented by Stetson et. al, National Syndromic Surveillance Conference poster, New York Academy of Medicine, October 2003) compared parking usage data from a parking facility servicing two hospitals, two large outpatient facilities and a medical school with counts of outpatient diagnoses for respiratory illness from military treatment facilities in the same geographic region covering a 18 month period ending June 30, 2003. The parking counts reflected only those users who paid per visit, and excluded those using passes to enter the parking facility. Students and employees of organizations served by the parking facility typically have monthly passes. Those paying per visit are patients, visitors and others having occasional business at the center.

The data covered a period in late February 2003 when many local area parents responded to alarming stories in the local print and electronic media describing a mysterious illness that seemed to be responsible for 5 unexpected child deaths, and advised seeking care for children who were ill. After investigation, public health officials found no unusual circumstances.

Data from both parking and clinical sources were smoothed by subtracting the present value from the average value of the eight preceding days of the same name (e.g. Tuesdays) and 95% confidence intervals were calculated for each day. The data streams were compared by inspection.

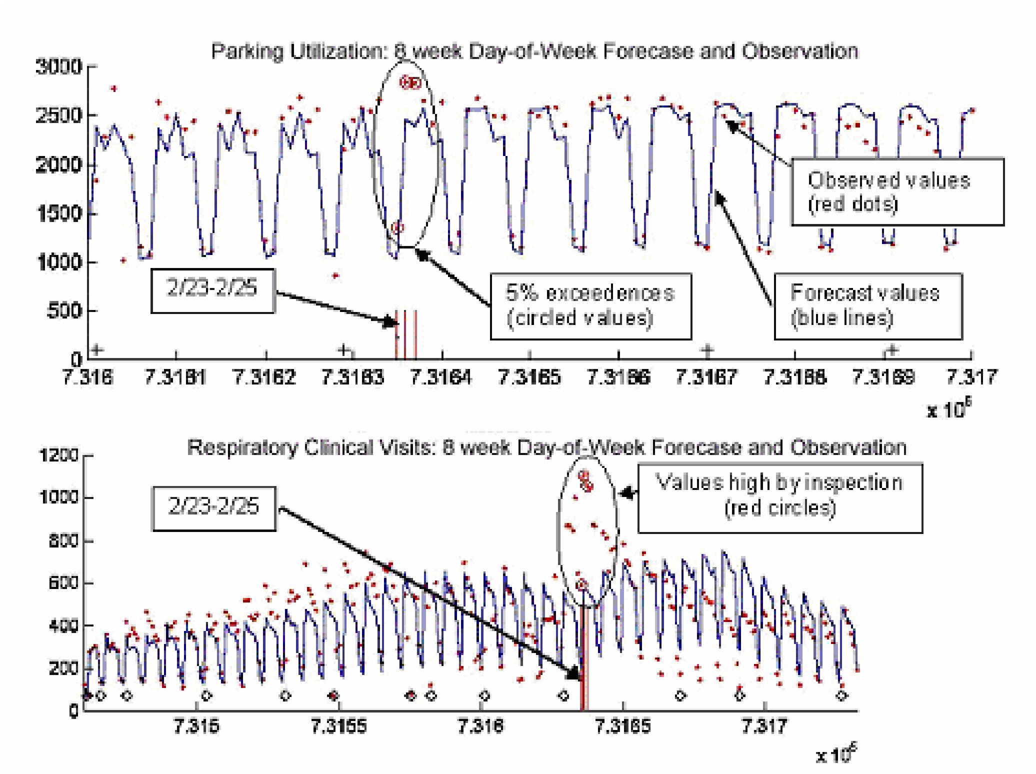

Figure 4.10 Parking Utilization (upper panel) and Respiratory Visits (lower panel) During Period Surrounding Media Event (2/23-2/25/2003) Showing Contemporaneous Peaks in Activity

4.11.3 Conclusions

The analysis illustrated what might be seen if a large part of the local population sought medical care because of an event that affected each individual at about the same time. The parking data closely mirrored an increase in respiratory medical visits during the time of area-wide concern about a possible unusual illness. The large excursions in both parking utilization and respiratory clinical diagnoses during the media event is unduplicated during the period of the analysis and appears to illustrate the sensitivity of the analysis to unusual events. It also documents the ability to monitor medical parking utilization as a way to trace increased medical visits. Such data may be timelier and more accessible with fewer privacy restrictions and cost than traditional clinical data. Collecting such data requires a rapidly reported method to monitor parking activity associated with medical facilities. Such circumstances may not be common.

4.12 Orthodontic Appointment Cancellations

Similar to absenteeism, tracking missed or cancelled appointments may provide an early indicator of illness in a community. Orthodontic appointments are usually scheduled well advance, and are made mostly for children. Cancellations are probably made with reluctance and may reflect unrelated illness

4.12.1 Data Characteristics

Orthodontist Appointment (PMOA) data is electronically collected by Orthodontic Centers of America, Inc for 14 orthodontist offices in Pennsylvania. These data are coded using a standard format that contains appointment date, appointment status, patient age, and patient zip code.

4.12.2 Experiments

Shah et al. evaluated the timeliness of PMOA for early detection of outbreaks when compared to hospital discharge data in Pennsylvania from 2000-2001 [55]. A missed appointment was defined as one that a patient cancelled, rescheduled, or did not show up. The investigators compared proportion of missed appointments (calculated by dividing the number of missed appointments by the total number (missed + kept) of appointments made) with a gold standard, complete ICD-9 coded hospital discharge data corresponding to lower respiratory tract infections due to pneumonia, influenza and bronchiolitis for the years 2000-2001. Patients were stratified as either under or above the age of eighteen years of age. Each time series was standardized and plotted against the date of admission to display seasonal outbreaks in the study population. The time latency between PMOA and hospital admission was measured by the cross-correlation function (CCF) (Figure 4.11). PMOA correlated with hospital admissions for respiratory illness but lags them by about 3 weeks for patients either above or below age 18 years.

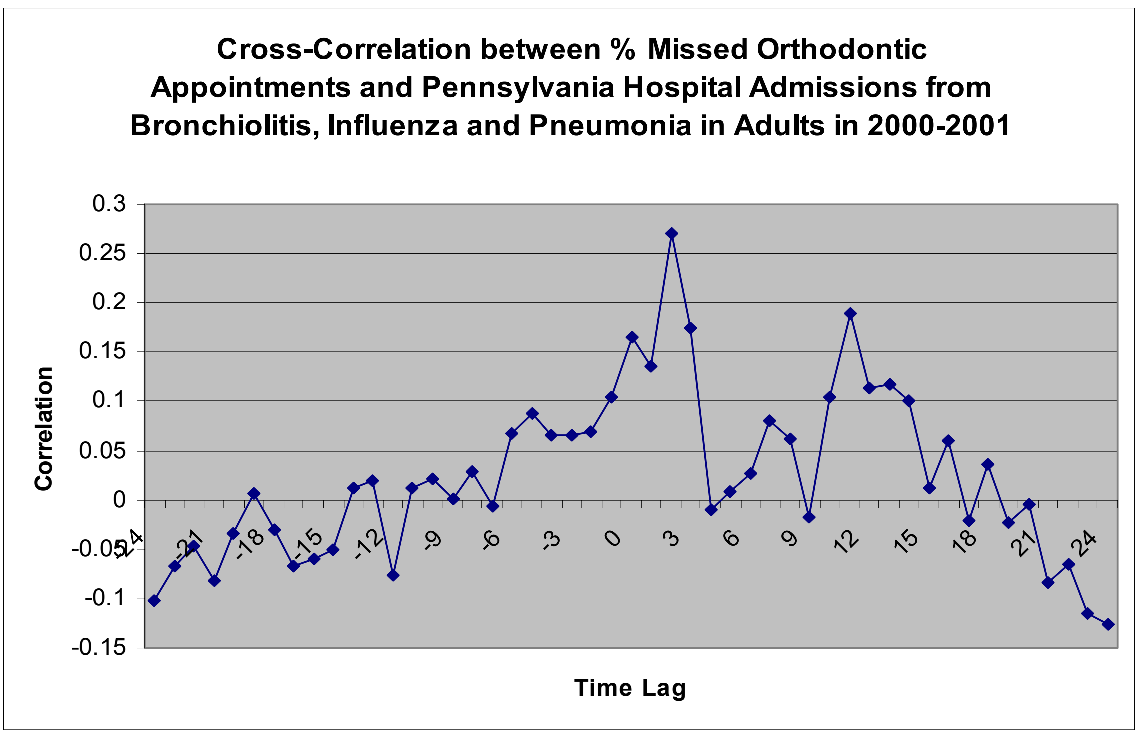

Figure 4.11 Cross correlation analysis: Missed Orthodontic Appointments and Hospital Admissions. This figure shows that the maximum correlation was 0.25 at a time lag of +3 weeks.

The reasons that this signal is not promising include: (1) only a tiny fraction of the population are scheduled to receive orthodontic care on a given day, so the sampling efficiency of this data source is low, (2) the background rate of cancellation is not small, so there is substantial noise, and (3) the data are non-specific containing no information about the cause for cancellation.

4.12.3 Conclusions

PMOA correlated with hospital admissions for respiratory illness but lags them by about 3 weeks for patients either above or below age 18 years. The investigators regarded the results as negative and feel that missed orthodontic appointments has little promise as a signal for community illness. Many factors may come to bear: small numbers of appointments per day, a relatively large cancellation rate, no reason for cancellation available, and demographic differences between orthodontic patients and the general population. Corrections for the day of week, time of year, and non-medical reasons for cancellation such as events, bad weather, and holidays may increase the value of the indicator. The timeliness of the signal may also be improved by using the date of cancellation in conjunction with the appointment date to, for example, count only cancellations that occur within a week of the appointment.

5 POTENTIAL DATA SOURCES

The BioALIRT program was terminated in its third of four years. In this section, we briefly describe data sources identified by the Bio-ALIRT contractors that remain to be acquired and or analyzed, including telephone call volume, cough detection, TV viewing, and agribusiness.

5.1 Telephone Call Volume

Telephone call volume from neighborhoods, apartment buildings, and schools— especially in the very early hours of the morning or late at night when levels are typically low--is a potentially cheap and early signal that requires further investigation. The BioALIRT contractors attempted without success to elicit participation from phone companies in the research (N.B. promising results were obtained for campus level phone traffic monitoring to medical facilities as described in Section 3).

5.2 Thermal Scanning

During the SARS outbreak of 2003, some airports and others used thermal scanners to detect incipient SARS cases among passengers. These scanners detected the elevated body temperatures that are often the first indicators of the start of an illness. Thermal scanners may be effective for detecting febrile illness before individuals have symptoms are severe enough to encourage absenteeism or the other sorts of detectable behaviors evaluated in the preceding sections. Serious issues involve assuring privacy rights of individuals, perhaps by using non-imaging scanners, or de-identifying images that are captured.

5.3 Cough Detection

Several investigators began exploration of cough detection, on the theory that there would be more background coughs in public spaces as seasonal or deliberate respiratory outbreaks unfold. Sound analysis algorithms are available that may be able to distinguish coughs from non-cough sounds in a classroom setting. Again, a significant issue is assuring personal privacy. One solution to privacy concerns is to field cough detectors that simply register that a cough has taken place, rather than supporting additional analyses, such as whether the increased coughs are all coming from one person. Field work will be required to demonstrate whether this source is a valuable early indicator of population illness.

5.4 TV Viewing

When people are sick enough to stay home they may change their television viewing habits. However, ratings for a local population are difficult to determine on an hourly or daily basis. An exception is pay-per-view utilization, which generates timely and detailed data streams. These data could include entertainment category information (adventure, sports) that might suggest the age of the viewer, and some geographic detail as well, such as user zip code. An additional, related source is video rentals. The data are collected electronically by point-of-sale systems, could contain detail such as entertainment category (children’s, adventure, drama, etc.) and may provide some geographic detail as well.

5.5 Agribusiness

Agribusiness has been suggested as an avenue for exploration. However, commercial animal populations are treated with numerous medications and may be inoculated against some of the zoonoses that initially make them seem attractive as early indicators of deliberate pathogen releases. BioALIRT had difficulty obtaining agribusiness datasets for the research, and the one dataset received did not have useful data (the level of aggregation of observations was too coarse).

5.6 EMS Usage

EMS systems commonly use electronic dispatch methods that collect data electronically. Data from EMS systems generally contain elements of time, user location and medical need associated with a call. BioALIRT contractors examined examples of this data but were not satisfied with the data quality of the samples available, and were also unable to acquire satisfactory contemporaneous gold-standard clinical data for correlation. Since EMS dispatch data is routinely collected in many locations, it may prove a readily available source for evaluation as a leading indicator of community illness.

6 References

1. Welliver, R.C., J.D. Cherry, K.M. Boyer, J.E. Deseda-Tous, P.J. Krause, J.P. Dudley, R.A. Murray, W. Wingert, J.G. Champion, and G. Freeman, *Sales of nonprescription cold remedies: a unique method of influenza surveillance.* Pediatr Res, 1979. **13**(9): p. 1015-7.
2. Wagner, M.M., R. Aryel, and V. Dato, *Availability and Comparative Value of Data Elements Required for an Effective Bioterrorism Detection System*. 2001, Agency for Healthcare Research and Quality: Washington DC. 184 pages.
3. Wagner, M.M., V. Dato, J.N. Dowling, and M. Allswede, *Representative threats for research in public health surveillance.* Journal of Biomedical Informatics, 2003. **36**(3): p 177-88.
4. Wagner, M.M., F.C. Tsui, J.U. Espino, V.M. Dato, D.F. Sittig, R.A. Caruana, L.F. McGinnis, D.W. Deerfield, M.J. Druzdzel, and D.B. Fridsma, *The emerging science of very early detection of disease outbreaks.* J Public Health Manag Pract, 2001. **7**(6): p. 51-9.
5. Hogan, W.R., F.-C. Tsui, O. Ivanov, P.H. Gesteland, S. Grannis, J.M. Overhage, N. Mikanatha, and M.M. Wagner, *Early detection of pediatric respiratory and diarrheal outbreaks from retail sales of electrolyte products.* J Am Med Inform Assoc, 2003. **10**(6): p. 555-62.
6. Chapman, W.W., L.M. Christensen, M.M. Wagner, P.J. Haug, O. Ivanov, J.N. Dowling, and R.T. Olszewski, *Classifying free-text triage chief complaints into syndromic categories with natural language processing.* AI in Medicine, 2004: in press.
7. Labrie, J., *Self-care in the new millenium: American attitudes towards maintaining personal health*. 2001, Consumer Healthcare Products Association: Washington, DC. p. 76.
8. McIsaac, W.J., N. Levine, and V. Goel, *Visits by adults to family physicians for the common cold.* J Fam Pract, 1998. **47**(5): p. 366-9.
9. Proctor, M.E., K.A. Blair, and J.P. Davis, *Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of Cryptosporidium infection.* Epidemiol Infect, 1998. **120**(1): p. 43-54.
10. Mac Kenzie, W.R., N.J. Hoxie, M.E. Proctor, M.S. Gradus, K.A. Blair, D.E. Peterson, J.J. Kazmierczak, D.G. Addiss, K.R. Fox, J.B. Rose, and J.P. Davis, *A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply.* N Engl J Med, 1994. **331**(3): p. 161-7.
11. Corso, P.S., M.H. Kramer, K.A. Blair, D.G. Addiss, J.P. Davis, and A.C. Haddix, *Cost of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin.* Emerg Infect Dis, 2003. **9**(4): p. 426-31.
12. Stirling, R., J. Aramini, A. Ellis, G. Lim, R. Meyers, M. Fleury, and D. Werker, *Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001.* Can Commun Dis Rep, 2001. **27**(22): p. 185-92.
13. Stirling, R., J. Aramini, A. Ellis, G. Lim, R. Meyers, M. Fleury, and D. Werker, *North Battleford, Saskatchewan, Spring 2001 waterborne Cryptosporidiosis outbreak*. 2001, Health Canada.

14. Wagner, M., J. Robinson, F. Tsui, J.U. Espino, and W. Hogan, *Design of a national retail data monitor for public health surveillance.* J Am Med Inform Assoc, 2003. **10**(5): p. 409-18.
15. Wagner, M., J. Espino, J. Hersh, W.R. Hogan, G. Shah, F.-C. Tsui, J. Hutman, G. Parks, C. Lewis, and R. Aller, *The National Retail Data Monitor for public health surveillance.* MMWR 2004 to appear.
16. Magruder, S. and E. Florio, *Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of public health.* Johns Hopkins University Applied Physics Laboratory Technical Digest, 2003. **24**(4): p. to appear.
17. Campbell, M., C.-S. Li, and et al, *An Evaluation of Over-The-Counter Medication Sales for Syndromic Surveillance*, in *IBM Research Technical Report*. 2004. p. in press.
18. Hassler, J., *Are they Swinging Together? A measure of linear comovement with an application to Swedish and foreign business cycles*. 1993, Institute for International Economic Studies, Stockholm University,: Stockholm, Sweden.
19. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*. 1965, New York, NY: McGraw-Hill.
20. Magruder, S., S. Happel Lewis, A. Najmi, and E. Florio, *Progress in Understanding and Using Over-the-Counter Pharmaceuticals for Syndromic Surveillance of Public Health.* special issue of MMWR, 2004: to appear.
21. Bendat and Piersol, *Engineering Applications of Correlation and Spectral Analysis*. 1993, New York, NY: John Wiley and Sons.
22. Florio, E., S. Magruder, S. Happel Lewis, and R. Wojcik. *Improving over-the-counter medication sales monitoring by modeling related factors*. in *Proceedings of the 131st Annual Meeting of the American Public Health Association*. 2003. San Francisco, CA.
23. Goldenberg, A., G. Shmueli, R.A. Caruana, and S.E. Fienberg, *Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5237-40.
24. Wallstrom, G.L., M.M. Wagner, and W.R. Hogan, *Using Surveillance Data from Actual Outbreaks to Characterize Detectability.* RODS Lab Tech Report.
25. Henry, J., S. Magruder, and M. Snyder, *Using Kaiser Permanente Nurse Advice Hotline Data for Syndromic Surveillance in the National Capital Area.* MMWR, 2004: to appear.
26. Platt, R., C. Bocchino, B. Caldwell, R. Harmon, K. Kleinman, R. Lazarus, A.F. Nelson, J.D. Nordin, and D.P. Ritzwoller, *Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program.* J Urban Health, 2003. **80**(2 Suppl 1): p. 25-31.
27. Espino, J.U., W. Hogan, and M.M. Wagner, *Telephone triage: A timely data source for surveillance of influenza-like diseases.* Proc AMIA Symp, 2003: p. 215-219.
28. Gesteland, P.H., *Unpublished results (see appendix).* 2003.
29. Chapman, W.W., J.U. Espino, J.N. Dowling, and M.M. Wagner, *Detection of multiple symptoms from chief complaints.* Technical Report, CBMI Report Series, 2003.
30. Espino, J. and M. Wagner, *The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness.* Proc AMIA Symp, 2001: p. 164-8.
31. Chapman, W., J.N. Dowling, and M. Wagner, *Syndromic Case Classification from Chief Complaints: a Retrospective Analysis of 527,228 Patients*. 2004, Technical Report, CBMI Report Series.

32. Ivanov, O., P.H. Gesteland, W. Hogan, M.B. Mundorff, and M.M. Wagner, *Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints.* Proc AMIA Symp, 2003: p. 318-22.
33. Chapman, W.W., J.N. Dowling, and M.M. Wagner, *Fever detection from free-text clinical records for biosurveillance.* J Biomed Inform, 2004. **37**(2): p. 120-7.
34. Tsui, F.C., M.M. Wagner, V. Dato, and C.C. Chang, *Value of ICD-9 coded chief complaints for detection of epidemics.* Proc AMIA Symp, 2001: p. 711-5.
35. Chapman, W. and P. Haug, *Comparing expert systems for identifying chest x-ray reports that support pneumonia.* Proc AMIA Symp, 1999: p. 216-20.
36. Jain, N.L., C.A. Knirsch, C. Friedman, and G. Hripcsak, *Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports.* Proceedings / AMIA Annual Fall Symposium, 1996: p. 542-6.
37. Knirsch, C., N. Jain, A. Pablos-Mendez, C. Friedman, and G. Hripcsak, *Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system.* Infect Control Hosp Epidemiol, 1998. **19**(2): p. 94-100.
38. Fiszman, M., W.W. Chapman, D. Aronsky, R.S. Evans, and P.J. Haug, *Automatic detection of acute bacterial pneumonia from chest X-ray reports.* J Am Med Inform Assoc, 2000. **7**(6): p. 593-604.
39. Chapman, W., M. Wagner, G. Cooper, P. Hanbury, B. Chapman, and L. Harrison, *Creating a text classifier to detect chest radiograph reports consistent with features of inhalational anthrax.* J Amer Med Inform Assoc., 2003. **10**(5): p. 494-503.
40. Li, C.-S. and C. Aggarwal. *Epi-SPIRE: A System for Environmental and Public Health Monitoring*. in *IEEE International Conference on Multimedia & Expo*. 2003.
41. Pipkin, A.B., *Anthrax*, in *Large Animal Internal Medicine, Second Edition*, B.P. Smith, Editor. 1996, Mosby-Year Book Inc: St. Louis, MO. p. 1246-1250.
42. Aiello, S., ed. *The Merck Veterinary Manual. Eighth Edition*. 1998, Merck and Co., Inc: Whitehouse Station, NJ.
43. Leib, M.S. and W.E. Monroe, eds. *Practical Small Animal Internal Medicine*. 1997, W. B. Saunders: Philadelphia.
44. Acha, P. and B. Szyfres, *Zoonoses and Communicable Diseases Common to Man and Animals. Second Edition*. 1987, Pan American Health Organization: Washington, DC.
45. Rohrbach, B., *Tularemia.* J. American Veterinary Medicine Association, 1988. **193**(4): p. 428-432.
46. Pavlin, J., P. Murdock, E. Elbert, C. Milliken, S. Hakre, J. Mansfield, and C. Hoge, *Conducting population behavioral health surveillance using automated diagnostic and pharmacy systems.* MMWR, 2004: p. in press.
47. Pew Internet and American Life Project, *Internet Health Resources: Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access.* 2003.
48. Tu, H.T. and J.L. Hargraves, *Seeking Health Care Information: Most Consumers Still on the Sidelines*, in *Issue Brief*. 2003, Center for Studying Health System Change: Washington, DC.
49. Johnson, H.A., M.M. Wagner, W. Hogan, W. Chapman, R. Olszewski, J.N. Dowling, and G. Barnas, *Analysis of Web Access Logs for Surveillance of Influenza*. Proceedings MEDINFO 2004, to appear.
50. Babin, S., *Weather and climate effects on disease background levels.* Johns Hopkins University Applied Physics Laboratory Technical Digest, 2003. **24**(4): p. in press.

51. Centers for Disease Control and Prevention, *Tropical Storm Allison rapid needs assessment - Houston, Texas, June 2001.* Morbidity and Mortality Weekly, 2002. **51**: p. 365-369.
52. Frischer, T.M., J. Kuehr, A. Pullwitt, R. Meinert, J. Forster, S. M, and H. Koren, *Ambient ozone causes upper airways inflammation in children.* American Review of Respiratory Diseases, 1993. **148**: p. 961-964.
53. Simpson, R.W., G. Williams, A. Petroeschevsky, G. Morgan, and S. Rutherford, *Associations between outdoor air pollution and daily mortality in Brisbane, Australia.* Archives of Environmental Health, 1997. **52**: p. 442-454.
54. Zeghnoun, A., P. Beaudeau, F. Carrat, V. Delmas, O. Boudhabhay, F. Gayon, D. Guincetre, and P. Czernichow, *Air pollution and respiratory drug sales in the city of Le Havre, France, 1993-1996.* Environmental Research, 1999: p. 224-230.
55. Shah, G., *Value of Missed Orthodontics Appointment for Timely Detection of Outbreaks*. 2003, RODS Laboratory: Pittsburgh, PA.

7 Appendix of Papers and Technical Reports

See appendix submitted with this report.