SYNDROMIC SURVEILLANCE FOR THE EARLY DETECTION OF INFLUENZA OUTBREAKS

by

Sara L. Rizzo

BS, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2005

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Sara L. Rizzo

It was defended on

September 29, 2005

and approved by

Joseph P.Costantino, Dr.PH
Professor and Director, NSABP
Biostatistics
Graduate School of Public Health
University of Pittsburgh

Vincent C. Arena, PhD
Associate Professor
Biostatistics
Graduate School of Public Health
University of Pittsburgh

Evelyn Talbott, Dr.PH
Professor
Epidemiology
Graduate School of Public Health
University of Pittsburgh

Joseph P.Costantino, Dr.PH

SYNDROMIC SURVEILLANCE FOR THE EARLY DETECTION OF INFLUENZA OUTBREAKS

Sara L. Rizzo, MS

University of Pittsburgh, 2005

Abstract

Syndromic surveillance is a new mechanism utilized to detect naturally occurring and bioterroristic outbreaks. The public health significance is its potential to alert public health to outbreaks earlier and allow a timelier public health response. It involves monitoring data that can be collected in near real-time to find anomalous data. Syndromic surveillance includes school and work absenteeism, over-the-counter drug sales, and hospital admissions data to name a few. This study is an assessment of an extension of the use of syndromic surveillance as an improvement to the traditional method to detect more routine public health problems, specifically, the detection of influenza outbreaks. The assessment involves the prediction of outbreaks in four areas during the period October 15, 2003 to March 31, 2004. The four areas studied included Allegheny County, Pennsylvania, Jefferson County, Kentucky, Los Angeles County, California, and Salt Lake County, Utah. Two aspects of community activity were used as the method for syndromic surveillance, over-the-counter pharmaceutical sales and hospital chief complaints. The over-the-counter sales encompassed a panel of six items including anti-diarrheal medication, anti-fever adult medication, anti-fever pediatric medication, cough and cold products, electrolytes, and thermometers. Additionally, two of the seven hospital chief complaints used in the RODS open source paradigm were monitored. These were constitutional and respiratory chief complaints.

Application of standard statistical algorithms showed that the system was able to identify unusual activity several weeks prior to the time when the local health departments were able to identify an outbreak using the standard methods. The largest improvement in detection using syndromic surveillance occurred in Los Angeles where the outbreak was detected 52 days before the Centers for Disease Control had declared widespread activity for the state. In each county over-the-counter sales detected the outbreak sooner then hospital chief complaints, but the hospital chief complaints detect the outbreaks consistently across the various algorithms.

More conclusive evidence regarding the possible improvement in outbreak detection with syndromic surveillance can be obtained once a longer time frame has passed to allow more historical data to accumulate. Conducting additional studies on influenza outbreaks in other jurisdictions would also be useful assessments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

In this era of increased concern for homeland security, there is a growing need for mechanisms of infectious disease surveillance to characterize and detect bioterroristic and naturally occurring outbreaks. It is important that these mechanisms be developed such that the identification of outbreaks is as timely as possible to allow a rapid response and containment of the spread of the disease.

Influenza surveillance is conducted by various organizations and strives to characterize influenza activity. Previous studies on syndromic surveillance and its ability to detect influenza have been conducted by various groups. A study in Southeastern Virginia in 2002 was able to detect influenza on January 14 utilizing ED data and the CuSum algorithm. Compared to the sentinel physicians reporting the peak not until January 23rd [1]. The New York City Department of Health and Mental Hygiene conducted a study in 2003 comparing a telephone survey of 2,433 individuals asking about flu-like illness, behaviors related to flu-like illness and diarrheal illness in the previous 30 days. The results of this survey were compared to the data collected by the NYC surveillance system over the same period. It was estimated that approximately 60 illnesses are represented for every visit to the ED [2]. Another study conducted in Belgium for the 1993 to 1994 influenza season reported increases of 100 percent in acute respiratory infections, 56 percent in work absenteeism, 26 percent of total pharmaceutical sales, and 14 percent in overall mortality during the peak of the influenza season [3].

This is a retrospective study of the influenza outbreak in four counties in the United States (Jefferson County, KY; Allegheny County, PA; Los Angeles County, CA; Salt Lake County, UT) for the 2003-2004 influenza season. The purpose of the study is to compare one of the traditional influenza surveillance methods, the State and Territorial Epidemiologist Report to syndromic surveillance and compare the date of detection of syndromic surveillance to the declaration of widespread activity.

2. Review of the Relevant Literature

In 1998 the Centers for Disease Control developed a plan for the early detection and control of disease outbreaks using syndromic surveillance [4]. This plan was accelerated with the intentional release of Bacillus anthracis in 2001 that resulted in an anthrax outbreak [5]. While syndromic surveillance was originally developed to detect large-scale release of biological agents, the current methodology reaches beyond this scope to help determine the size, spread, and temporal aspects of an outbreak after it is detected. It can also provide reassurance that an outbreak is not occurring.

Additionally in 2000 the US Department of Health and Human services conducted an extensive review of the existing capabilities to detect four major public health threats, specifically emerging infectious disease, antimicrobial threats, bioterrorism, and pandemic influenza. The conclusion was that funding needs to be supplied for efficient, "easy to use", and rapid automated reporting [6]. This can take two forms. The first is facilitating the clinicians active reporting, or the use of syndromic surveillance [7].

Currently there are two forms of outbreak surveillance disease surveillance and syndromic surveillance. Disease surveillance is accomplished by tracking disease incidence. It is a routine practice of public health agencies to report to the CDC and the state conditions which may evolve into epidemic proportions from the usual endemic levels. Many states including PA utilize the NEDDS system (National Electronic Disease Surveillance System). NEDSS standards facilitate information exchange. Conditions such as number of children with elevated blood lead levels, hepatitis, viral and bacterial meningitis, influenza, and other childhood infectious diseases are reported by doctor's offices and county labs on a routine basis. Currently, influenza is not a required reportable disease by physicians. Laboratory confirmed cases are required to be reported. These are disease driven and are the current gold standard in determining when an "epidemic" is evolving at the state or national level.

NEDSS has four main initiatives. They are: 1) to detect outbreaks rapidly and to monitor the health of the nation; 2) facilitate the electronic transfer of appropriate information from clinical information systems in the health care system to public health departments; 3) reduce provider burden in the provision of information; and 4) enhance both the timeliness and quality of information provided [8]. The sensitivity and specificity of this system is directly affected by its technical abilities, and can be hampered by incomplete data reporting. These technical abilities allow for a faster data transmission [9].

Trying to detect bioterroristic agents from the symptoms that are present requires an astute clinician to make this conclusion from various individuals that are diagnosed with a particular disease [10,11]. The symptoms from the release of inhalation anthrax, for example will show increases of influenza-like illness, but generally clinicians diagnose individuals rather then epidemics [12,13].

Syndromic surveillance relies on protocols or automated routines for the collection of data. This ensures near real-time data acquisition [12]. There are various sources of syndromic surveillance. The sources range from emergency room chief complaints, ambulance dispatch data, and clinical diagnosis data [14-18]. Systems also exist that incorporate emergency room syndromes, private practice billing codes grouped into syndromes, and veterinary syndromes [19] Calls to poison control centers [20], nurse help-line logs [21] , and absenteeism to schools [17] are also data streams collected for syndromic surveillance purposes.

Syndromic surveillance strives to identify disease clusters earlier then traditional surveillance methods and to assist in the initiation of a rapid response [22]. An example of an outbreak that may have been detected more readily by syndromic surveillance than by traditional disease surveillance occurred among Milwaukee-area residents in 1993. It was a water-borne outbreak resulting in diarrhea among 400,000 people. This epidemic was only detected after shortages of anti-diarrheal medicine and enteric culture media were reported. If syndromic surveillance would have been employed in the Milwaukee area at the time of the epidemic, a large spike in anti-diarrheal sales may have been identified many days before the supply of the medication was depleted allowing for an earlier detection of the outbreak and an earlier initiation of public health response [23].

The public health department in New South Wales employed a surveillance system for the 2000 Olympic Games in Sydney. Seven major components made up the surveillance system. They were: 1) enhanced surveillance of communicable diseases; 2) sentinel emergency departments; 3) medical encounters at Olympic venues; 4) cruise ship surveillance; 5) Olympic venue food safety and health environmental inspections; 6) bioterrorism; and 7) global epidemic intelligence. Each of these components provided data at least on a daily basis to permit identification of an issue within 24 hours. The Olympic Games did not have any major health impacts or outbreaks, but was well suited to detect them if they presented [24].

Influenza has profound health effects on the population causing an estimated 20,000 deaths and 200,000 hospitalizations yearly [25, 26]. The costs associated with influenza exceed $12 billion every year as well [26].

One study conducted at the Memorial Medical Center, a community teaching hospital for Southern Illinois University School of Medicine, investigated the impact of rapid reporting of respiratory viruses and its financial and health effects. The study was conducted over two years. The second year of the study rapid reporting of positive cultures to the clinicians was initiated. The second year of the study produced decreased mortality, length of stay, hospital costs, and better antibiotic stewardship [27]. A study in Hong Kong also showed clinical and financial impacts of virology data in the pediatric population [28].

Many methods have been published to support timely detection and monitoring through a specified information structure [29-33] and these methods are emerging rapidly. Algorithm development for these data is also evolving. Algorithms and methods used in syndromic surveillance include regression algorithms for the diagnosis of genital ulcer disease [34], multiple temporal and spatio-temporal outbreak-detection algorithms [35], algorithms for the evaluation of statistical detection of peaks [36], and recursive least squares adaptive filter, an autoregressive linear model [37] to name a few.

ICD-9-Coded Chief Complaints have also been assessed for their sensitivity and specificity for detecting influenza. A study from December 5, 1999 to December 2, 2000 was conducted utilizing ICD-9 codes that fell into either a respiratory or influenza set. The Serfling method was utilized to predict the beginning of the influenza outbreak. This was compared to the pneumonia and influenza deaths prediction, the current gold standard. The study resulted in a sensitivity of 100 percent and a positive predictive value of 50 percent for the respiratory set and 25 percent for the influenza set. The ICD-9 codes were also able to detect influenza one week prior to the gold standard [ICD9 codes 38].

The Real-time Outbreak and Disease Surveillance (RODS) Laboratory [39] is a collaboration between Dr. Michael Wagner and colleagues at the University of Pittsburgh Center for Bioinformatics and the Auton Lab at the Carnegie Mellon University School of Computer Science. The laboratory was founded in 1999 to investigate methods for real-time detection and assessment of disease outbreaks. The focus of the project is algorithm development, assessment of novel types of surveillance data, natural language processing and analyses of syndrome detectability. The laboratory is home to four large projects that work with health departments to create surveillance systems: RODS software development, the Public Health Data Center, the National Retail Data Monitor (NRDM) and the BioWatch Support Program.

The open-source phase of the project began in August 2003, when the University of Pittsburgh released the source code for RODS under the GNU General Public License (GPL). The RODS Open Source Project is a collaboration involving academia, open source developers, health departments, hospitals and medical centers, foundations, and industries whose objective is to rapidly increase the level of deployment of syndromic and potentially other surveillance systems. At the same time, the RODS Laboratory created a web site [40] to distribute the source code and to allow and encourage developers, consultants, academics, and companies to participate in the further development and modification of the software. RODS is written in JAVA as a set of software modules (using JDK 1.4 and J2EE.) As a modular system, a subset of RODS modules can be used within existing public health surveillance projects or all of the modules can be used to create an end-to-end disease outbreak and surveillance system. Turnkey software packages, hardware prerequisites, software prerequisites and source code are available from the website.

RODS is readily available NEDSS and PHIN-compliant software for building public health surveillance systems. The RODS Laboratory has designed an application that provides a real-time syndromic surveillance system to a health department without the large costs typically incurred with the design and implementation of such a system.

The RODS system is syndromic driven. The complaints are not as specific but have utility when coupled with other laboratory and clinical information. The seven complaints are diarrhea, respiratory, constitutional, hemorrhagic, gastrointestinal, rash, and neurological.

The chief complaint for every emergency room hospital admission is available in databases in the form of chief complaints provided by patients visiting the emergency department. These chief complaints are then placed into one of the seven syndromic categories mentioned above by a Bayesian text classifier. Each chief complaint is placed into the syndrome of which it has the highest probability of being. This classifier was tested on 800 chief complaints in Utah and produced areas under the receiver operating curve (ROC) between 0.95 and 1 [41]. The sensitivity and specificity were also calculated and produced results of 52 and 89 percent, respectively [42].

The New York City Department of Health and Mental Hygiene (DOHMH) institute a syndrome coding system in New York [43]. This coding system was compared to the RODS Bayesian text classifier Complaint coder (CoCo) to determine if the two agreed on free-text encoding and syndrome diagnosis. The results did show overall agreement between the two, but also a need for consensus in classifying the free-text classifiers [44].

The National Retail Data Monitor utilizes over-the-counter drug sales obtained from Universal Product Codes (UPC) [45]. OTC data sources are useful since the first response that most patients have when they develop an illness is to obtain treatment with such medications.

This has been proven by two surveys. The first was conducted by the Consumer Healthcare Products Association in 2001. This was a random digit dialing survey of 1,505 individuals. When participants were asked what their response was to flu-like and cold symptoms 72 percent treated themselves with OTC products, and 42 percent stated that this was their first response. Thirty-four percent responded with self-observation first and only 9 percent reported seeking professional medical care as their first response. The actions that patients took for headache symptoms was more astounding with 81 percent's first action being self-administered OTC products, while 52 percent responded with self-observation and only four percent sought medical attention first [46]. The second survey was a population-based survey in Ontario, Canada with 42,333 adult participants. In this study 76 percent responded with OTC medication as their first response while only 14 percent sought medical treatment for upper-respiratory tract infections [47].

Collecting OTC sales data also has interesting characteristics. First, the sales data is collected by the utilization of UPC codes. UPC codes are barcodes placed on products that enable the sale of the product to be scanned and the data collected in real time. UPC codes are able to distinguish between each item. For example, two of the exact same items that have different quantities (i.e. 4 oz. vs. 8oz.) have different barcodes to distinguish between each other. Secondly, the market share of OTC products is held by a small number of national companies. Five national retailers make up 48 percent of the market share of OTC products. The top 10 OTC retailers account for 65 percent of the market share and the top 20 account for 76 percent, and lastly, these data can also be captured with a relatively low cost [45].

3. Methods

There are three phases of activity involved in this assessment. The first was the data collection phase. This involves the collection of OTC sales data, ED chief complaint data as well as information on the influenza outbreak. The second phase was the prospective analysis of data; and the third phase was the retrospective analysis of data. These three steps are discussed in further detail in the sections below.

3.1. Data Collection

Data used for in the analysis came from three sources. The first source was the OTC drug sales data collected by the National Retail Data Monitor (NRDM). The second was emergency room chief complaint data collected by the RODS system, and the third was the public health information available which describes the outbreak. The nature of the information collected from each of these sources is discussed in the following three sessions.

3.1.1. National Retail Data Monitor

The Real-time Outbreak and Disease Surveillance (RODS) Laboratory at the University of Pittsburgh maintains the National Retail Data Monitor (NRDM) which collects OTC drug sales data from retailers UPC codes across the country. The OTC data are collected in near real-time and are placed into one of eighteen categories. These sales data are then aggregated and can be viewed on the state, county or zip code level.

The NRDM information used in this assessment included data for over 20,000 stores from food, drug and mass merchandising chains. Further details about the nature of the retailers that provided the OTC data can not be disclosed. Due to agreements that are established with the retailers, the identity of the retailers must remain confidential. This is a safeguard to prevent competitors from learning of the market share of the participating retailers within the industry.

3.1.2. Emergency Room Data

In addition to the data from the NRDM the RODS system also includes ED hospital admission chief complaint data that was used in this study. The data includes information for, hospitals in four states (PA, UT, NJ, OH). A free-text classifier is used to categorize each chief complaint [41] in the participating hospitals into one of seven categories. These data are available to be analyzed on a state, county, or zip code level.

Both the OTC sales and ED hospital admissions data were downloaded from a user interface at the RODS site. Public health professionals can access this site upon signing a user agreement

3.1.3. Data Collection

Emergency department chief complaint data is sent via the Health Level 7 (HL7) protocol from participating hospitals in real-time. One a clinical encounter occurs this data is placed into the hospitals system and is transmitted to RODS. The data is automatically classified into one of seven syndromes using Bayesian classifiers.

OTC data is transmitted in a similar way except the OTC category in which the data is placed is based on the UPC code. This reporting is received in a batch mode on a daily basis [48].

3.1.4. Public Health Information

Several aspects of public health information are important for this assessment. This data were gathered through press releases from state and local Departments of Public Health and from media reports. The public health information that was collected for this study included the outbreak location, the organism causing the outbreak, demographics of the affected population, epidemic curves and reference dates of occurrence. The determinations of the reference dates for the outbreaks are described in the next paragraph. Using this information, a chronology was developed. The chronology highlights public health surveillance activity as well as the syndromic surveillance efforts that existed in the affected community throughout the outbreak.

The State and Territorial Epidemiologists Report (STER) generated by the CDC was used to select the reference date. The reference date used was that when widespread activity was declared in each state. This was chosen as it is a good indicator of the beginning of the influenza outbreak and it provides a date that is a consistent measure in each outbreak. The STER report is generated by the CDC on a weekly basis and reports influenza activity by geographic area. The information is reported through the PHLIS. There are five levels of influenza activity that are reported. They are no activity, sporadic, local, regional and widespread. No activity indicates that there were no laboratory confirmed cases and no reported increases in the number of influenza-like-illness reports. Sporadic activity indicates that a small number of laboratory-confirmed influenza cases or a single influenza outbreak had been reported, but there was no increase in cases of influenza-like-illness (ILI). Local activity is reported when outbreaks of influenza or increases in ILI cases do exist but within only one region of the state. Regional activity is reported when outbreaks of influenza, increases in ILI or laboratory confirmed cases are recorded in at least two, but less then half of the regions of the state. Widespread activity is reported when Outbreaks of influenza or increases in ILI cases and recent laboratory-confirmed influenza occurred in at least half the regions of the state

3.2. Prospective Analysis

The algorithm used for the prospective analysis was that developed by Zhang et al. this algorithm is designed to detect outbreaks from time series data using wavelet transform [49]. This algorithm was run on a prospective basis and its output saved in a database for historical viewing. Signals are generated at three standard deviations.

3.3. Retrospective Study

Seven different algorithms were used for the retrospective study. There were four variations of the CuSum-EWMA algorithm, two variations of an ARIMA algorithm, and the wavelet transform.

3.3.1 CuSum-EWMA

Two variants of the MatLab CuSum function were used employing the methodology described by Stoto et al [50]. There were two variations of the algorithm. These variations adjusted the weight used in the EWMA calculation, and the number of days used for the standard deviation calculations.

3.3.2 ARIMA

Two ARIMA models were used; both utilize the SAS proc ARIMA command. The first was a fixed parameter model. As described by Reis et al. [51] it uses auto-regressive and moving average parameters of one. The second model was adaptive and selected the best fitting parameters for executing the ARIMA forecast.

3.3.3 WAVELET

The wavelet algorithm was that of Zhang et al [49] referenced in the prospective analysis that utilizes the wavelet MatLab function, except this algorithm was run retrospectively. Differences in the outcomes of these two analyses may have arisen if data from a retailer was delayed. This could be due to technical problems on the hospital or retailers end that delay the reporting process. In addition the prospective wavelet algorithm sent a signal at three standard deviations. The retrospective algorithm sent signals based on a empirical false alarm rate based on historical data. Differences in the outcome of the prospective and retrospective analysis are the result of differing false alarm rates and incomplete data.

3.4. Timeliness

The timeliness of detection is the primary end point. This is the difference between the reference date and the date of the signal. This parameter is calculated for each data stream in each algorithm.

Differing false alarm rates will be used to establish the tradeoff between timeliness and the number of false alarm rates per year. Timeliness calculations are calculated for false alarm rates from two to twelve per year. Activity monitoring operator characteristic (AMOC) curves will be used to display this tradeoff.

4. Results

The results of four influenza outbreaks included in this study are presented below. The first two influenza outbreaks occurred in Salt Lake County, Utah and Allegheny County, Pennsylvania. Both OTC data chief complaint data were available for these outbreaks. The other two outbreaks occurred in Jefferson County, Kentucky and Los Angeles County, California. Only OTC data were available for these outbreaks as the RODS system was not enacted in these jurisdictions at the time of the outbreaks.

For the prospective analysis, the wavelet algorithm was run on only six OTC categories due to data storage limitations. Five of which were utilized for this analysis. They are anti-diarrheals, anti-fever pediatric (APP), cough and cold products, electrolytes, and thermometers.
Six OTC categories and two chief complaint categories were investigated including anti-diarrheals, APP, anti-fever adult (APA), cough and cold products, electrolytes, thermometers and constitutional and respiratory chief complaints.

Signals were generated for false alarm rates between two and twelve per year. The results of the signals produced by false alarm rates of two and four per year are compared.

4.1. Outbreak 1: Salt Lake County, Utah

On September 8, 2003 the first positive rapid antigen test was reported in Salt Lake County. The public health department did not respond to this since it was a single lab confirmed isolate. In late October 2003 there were also few Influenza-like illness reports to the public health department that were ignored. Widespread activity was not reported until the week ending of November 22, 2003.

The times series curves for Salt Lake County, Utah are displayed in Figures 1 through 8. The time series in each figure displays data from September 1, 2003 to March 32, 2004.

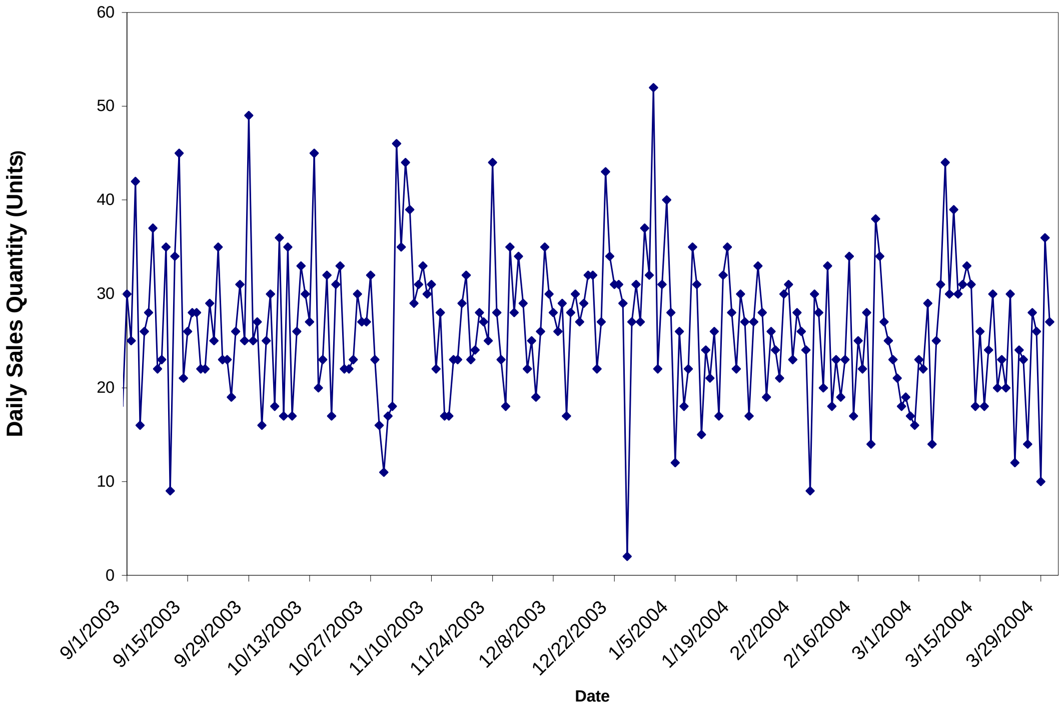

Figure 1 Anti-Diarrheal Sales Salt Lake County August 1, 2003 to March 31, 2004

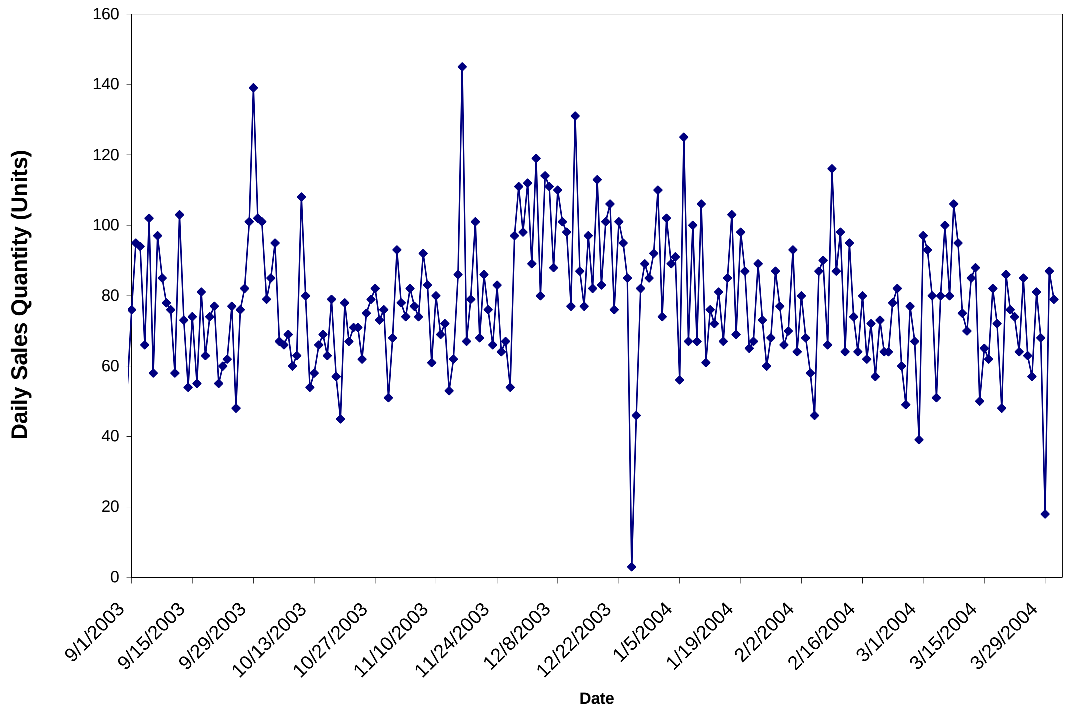

Figure 2 APA sales Salt Lake County August 1, 2003 to March 31, 2004

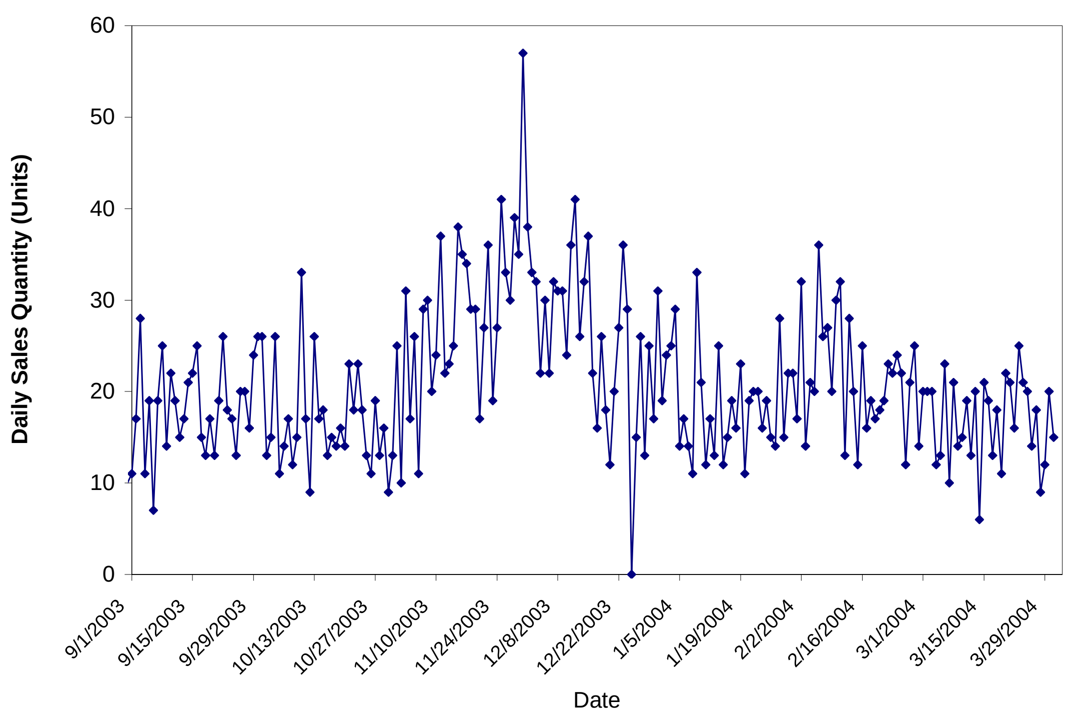

Figure 3 APP sales Salt Lake County August 1, 2003 to March 31, 2004

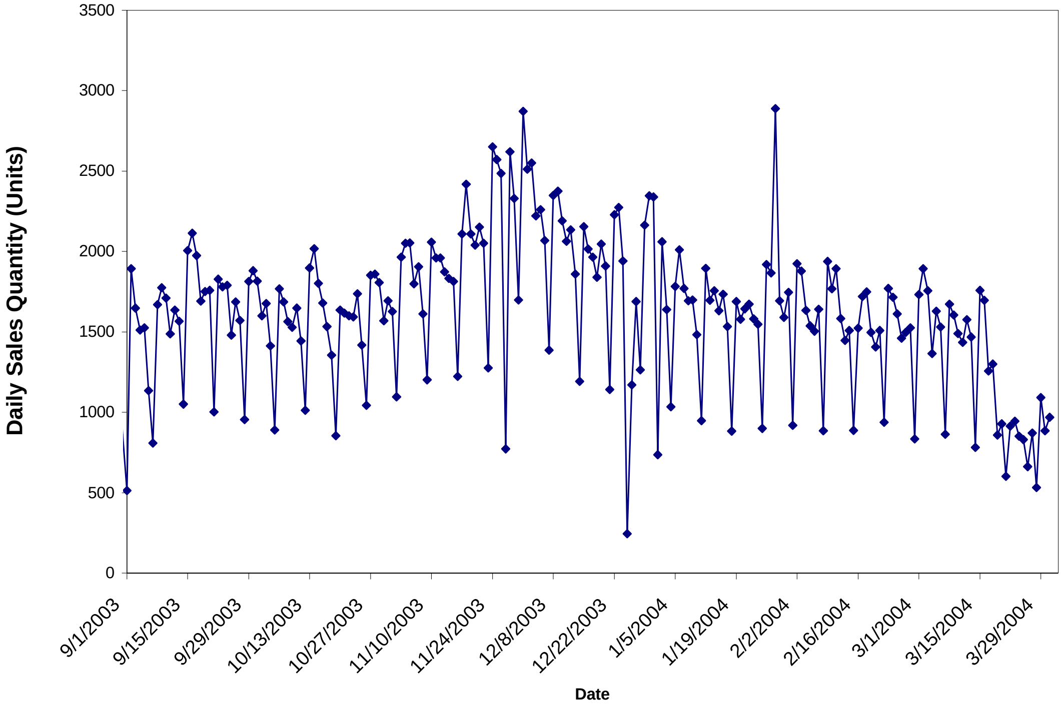

Figure 4 Cough/Cold sales Salt Lake County August 1, 2003 to March 31, 2004

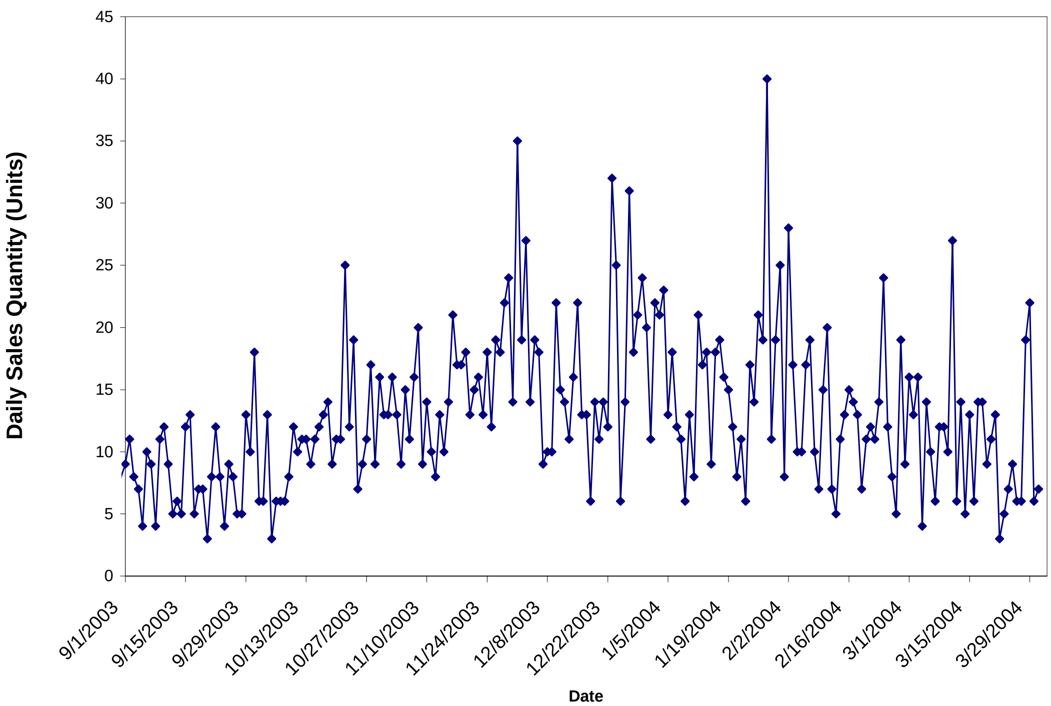

Figure 5 Electrolytes sales Salt Lake County August 1, 2003 to March 31, 2004

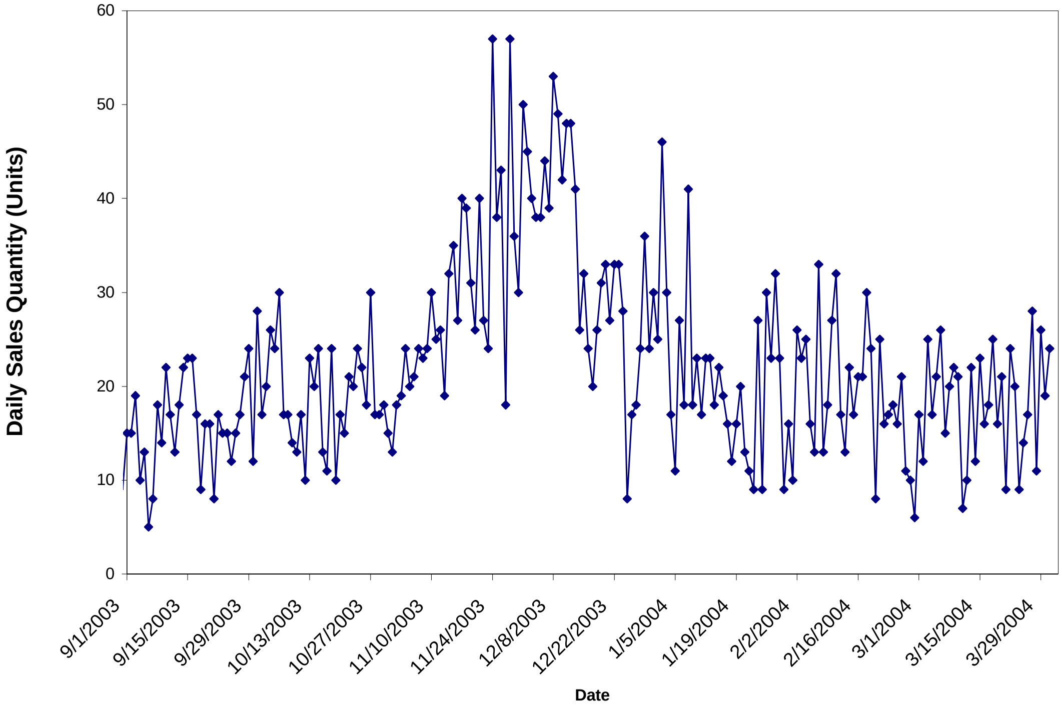

Figure 6 Thermometer Sales Salt Lake County August 1, 2003 to March 31, 2004

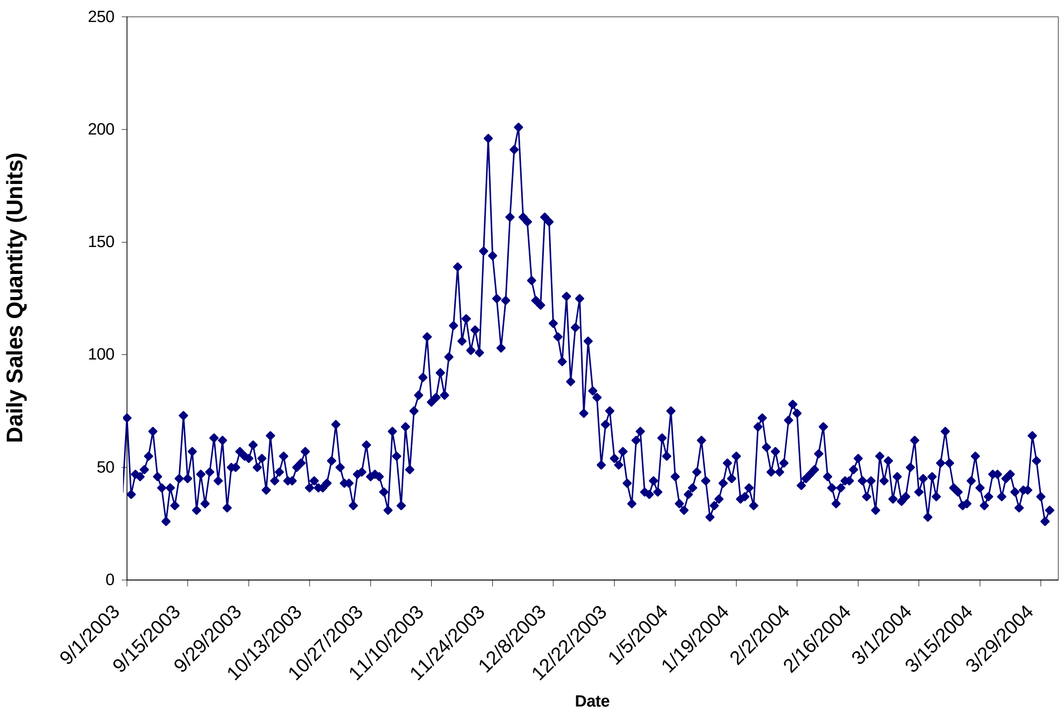

Figure 7 Constitutional Chief Complaints Salt Lake County August 1, 2003 to March 31, 2004

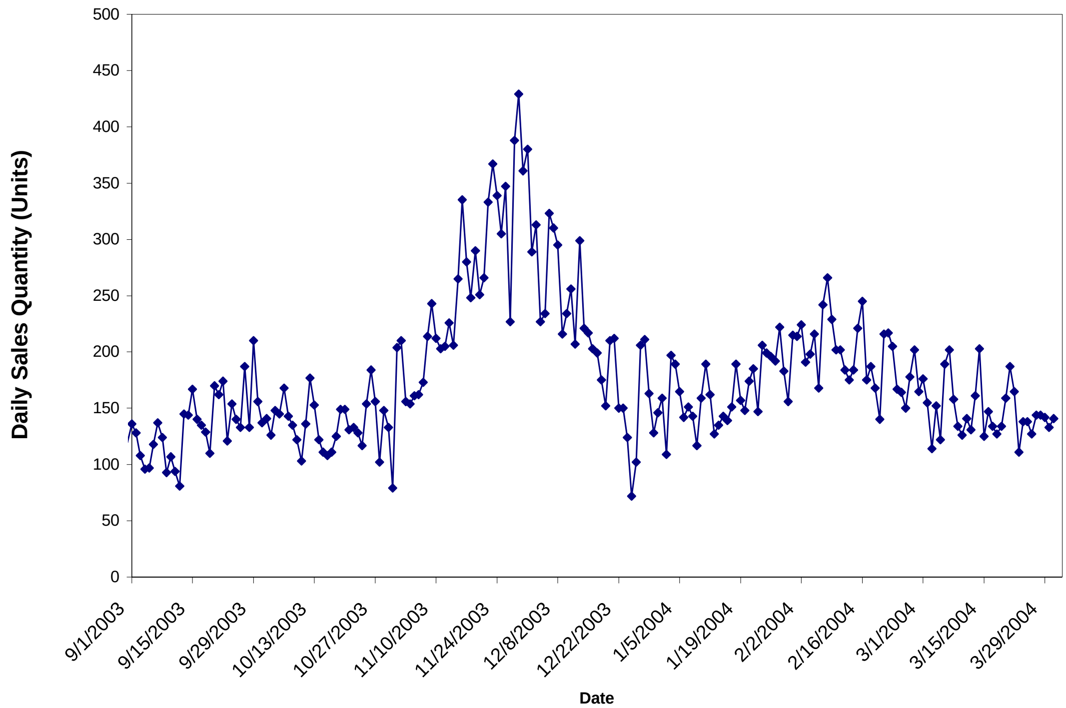

Figure 8 Respiratory Chief Complaints Salt Lake County August 1, 2003 to March 31, 2004

Two prospective alarms signaled for Salt Lake County. The first was in thermometer sales which signaled on November 24, 2003 with a standard deviation of 3.67. The second alarm signaled on November 30, 2003 in the anti-fever pediatric category with 3.07 standard deviations. The anti-diarrheals, cough/cold, and electrolytes categories did not produce any prospective signal. These prospective results are displayed in Table 1.

Table 1 Prospective Signals for Salt Lake County, Utah

OTC	Date of signal	Location	Standard Deviations	# of signals for the 2 months prior to the outbreak	# of signals for the 2 months after the outbreak
Thermometers	11/24/2003	Salt Lake	3.67	0	0
APP	11/30/2003	Salt Lake	3.07	0	0

The earliest alarm generated from OTC data in the retrospective analysis for a false alarm rate of two per year occurred on October 22, 2003 in the cough and cold and electrolytes data streams. The earliest alarm for a false alarm rate of four per year was in the electrolytes data stream occurring on October 22, 2003 as well. The electrolytes data stream signaled consistently for false alarm rates of both two and four per year in every algorithm. Each signal was sent on October 22, 2003 with corresponding timeliness of -31 days. The other OTC categories had more variation in their timeliness. For example, the earliest signal for a false alarm rate of two per year in cough and cold category was on November 5, 2003 with the CuSum algorithm with a weight of 0.05 and a window of 10 for the standard deviation. The latest signal was on January 28, 2004 in the CuSum algorithm with a weight of 0.20 and an infinite window.

Neither of the ED chief complaints signaled in any of the CuSum algorithms for a false alarm rate of two per year. For a false alarm rate of four per year the earliest constitutional alarm was generated on October 21, 2003, thirty-two days before widespread activity, the latest alarm occurred on November 8, 2003, fourteen days prior to widespread activity. The earliest respiratory alarm occurred on November 1, 2003 and the latest alarm occurred on November 17, 2003. The retrospective analysis results are displayed in Table 2.

Table 2 Retrospective Signals for Salt Lake County, Utah

	CuSum								Wavelet		ARIMA			
	weight = 0.05 window = infinite		weight = 0.20 window = 10		weight = 0.20 window = infinite		weight = 0.05 window = 10		Wavelet		Adaptive		ARIMA (1,0,1)	
	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Anti-diarrheals	**	**	**	-18	-18	-19	**	**	**	**	**	-20	**	-20
APA	14	13	-6	-6	**	**	-10	-16	**	**	-6	-6	-6	-6
APP	-11	-14	**	-5	**	-17	-5	-7	8	8	8	-11	8	-11
Cough & Cold	-2	-3	67	4	3	-17	-17	-17	2	2	12	12	-5	-5
Electrolytes	-31	-31	-31	-31	-31	-31	-31	-31	-31	-31	-31	-31	-31	-31
Thermometers	-10	-12	-7	-12	**	-7	-12	-12	2	2	2	2	2	-5
Constitutional	**	-14	**	-32	**	-14	**	**	-14	-14	-15	-15	-16	-16

4.2. Outbreak 2: Allegheny County, Pennsylvania

The first two cases of influenza were identified in Allegheny County in the week ending on November 11, 2003. As of December 3, 2003, 22 cases of influenza had been identified. This is the highest number of confirmed case for this time period since the Allegheny County Influenza Surveillance Program began in 1991. The outbreak was declared over on March 17, 2003. Widespread activity was declared on November 22, 2003.

The time series curves for Allegheny County, Pennsylvania are displayed for the time period of September 1, 2003 to March 31, 2004 in figures 9 through 16.

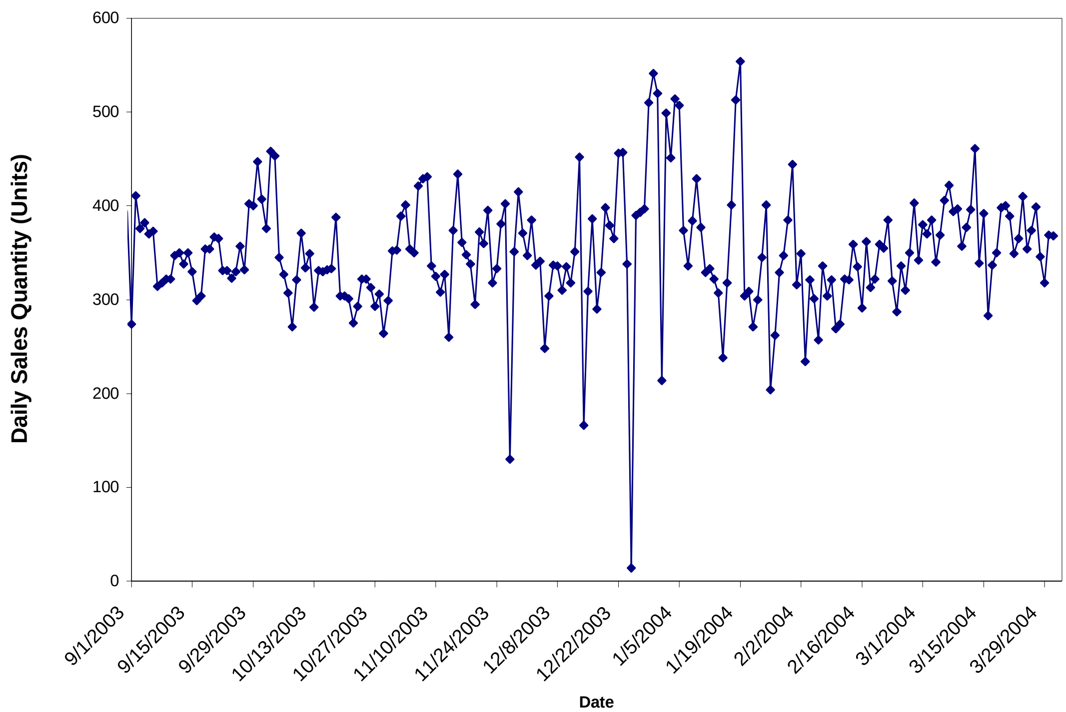

Figure 9 Anti-Diarrheal sales Allegheny County August 1, 2003 to March 31, 2004

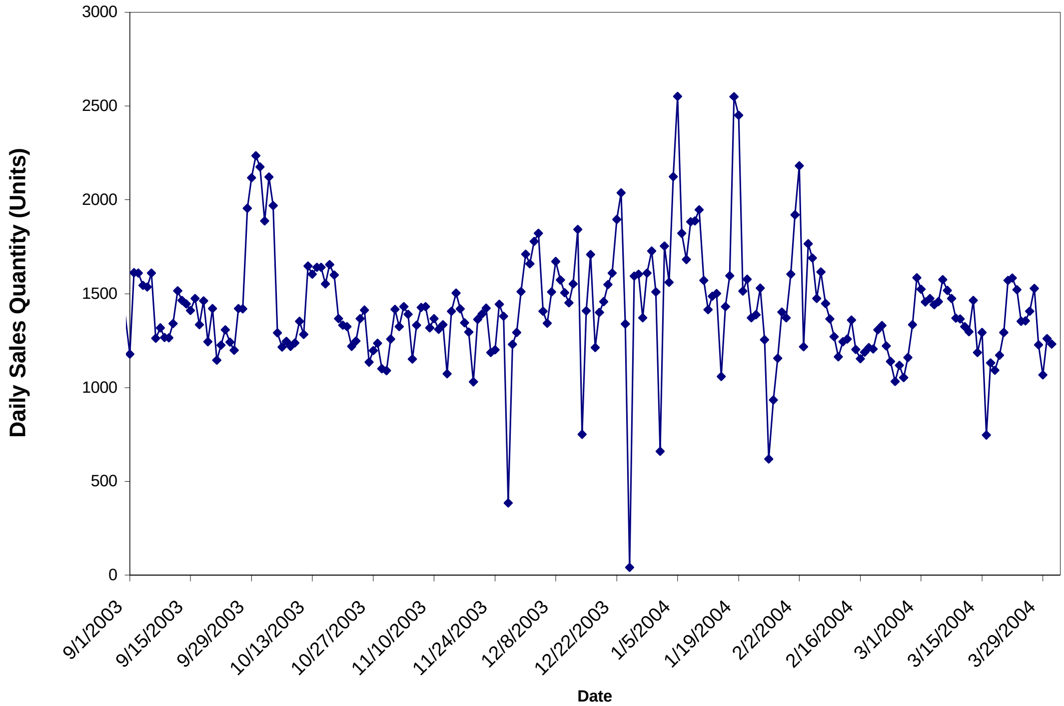

Figure 10 APA sales Allegheny County August 1, 2003 to March 31, 2004

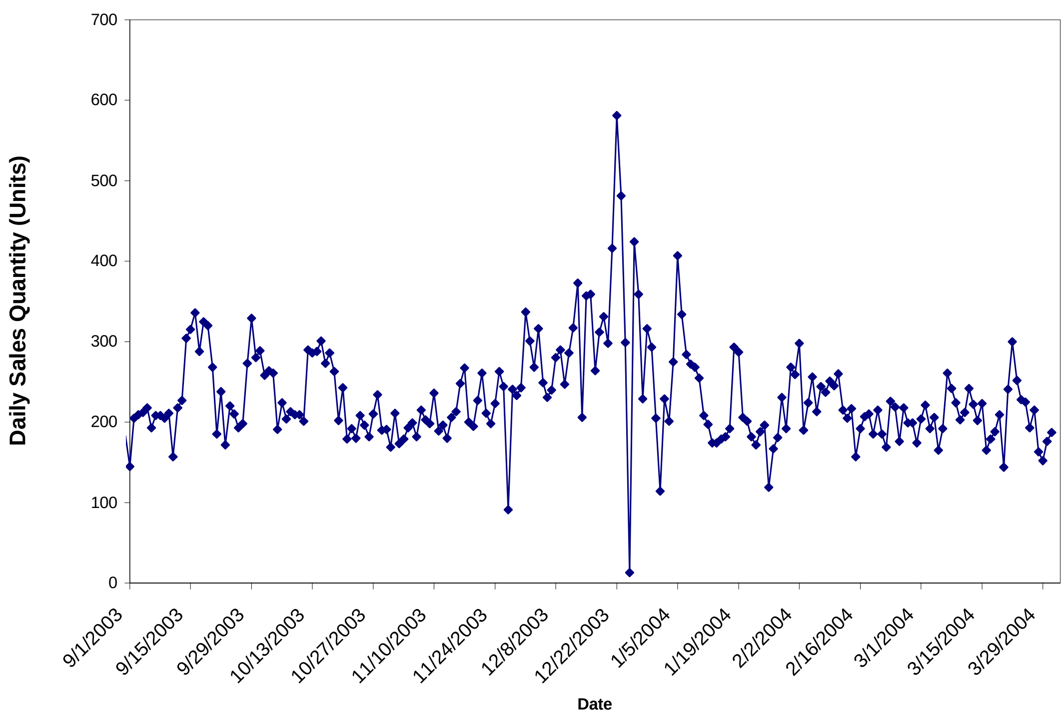

Figure 11 APP sales Allegheny County August 1, 2003 to March 31, 2004

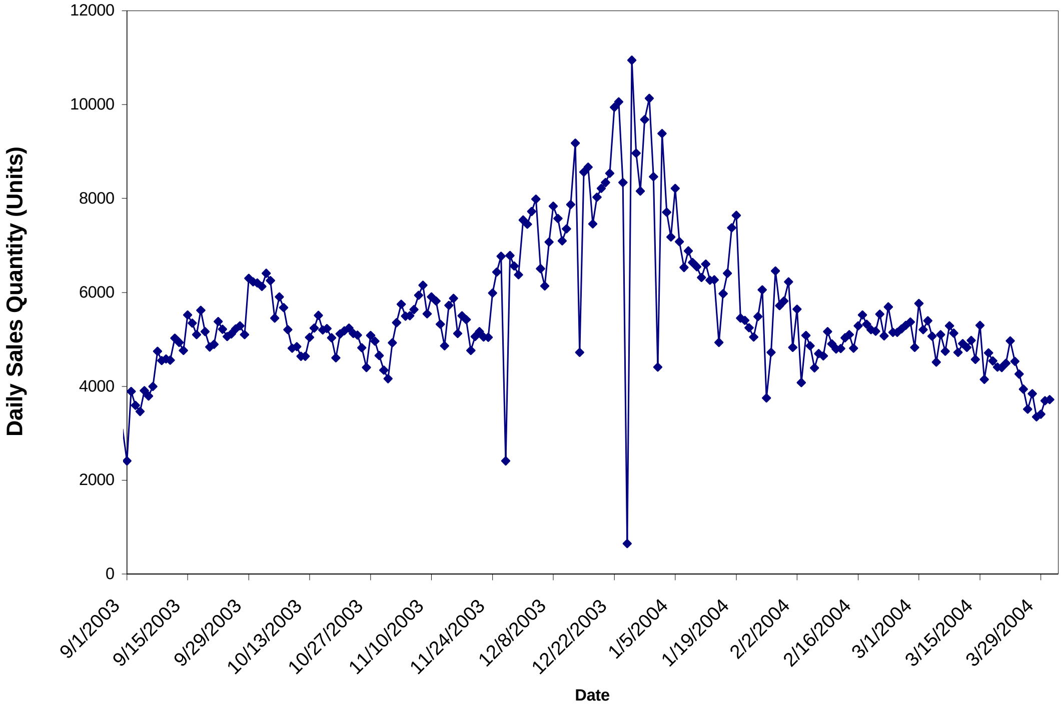

Figure 12 Cough/Cold sales Allegheny County August 1, 2003 to March 31, 2004

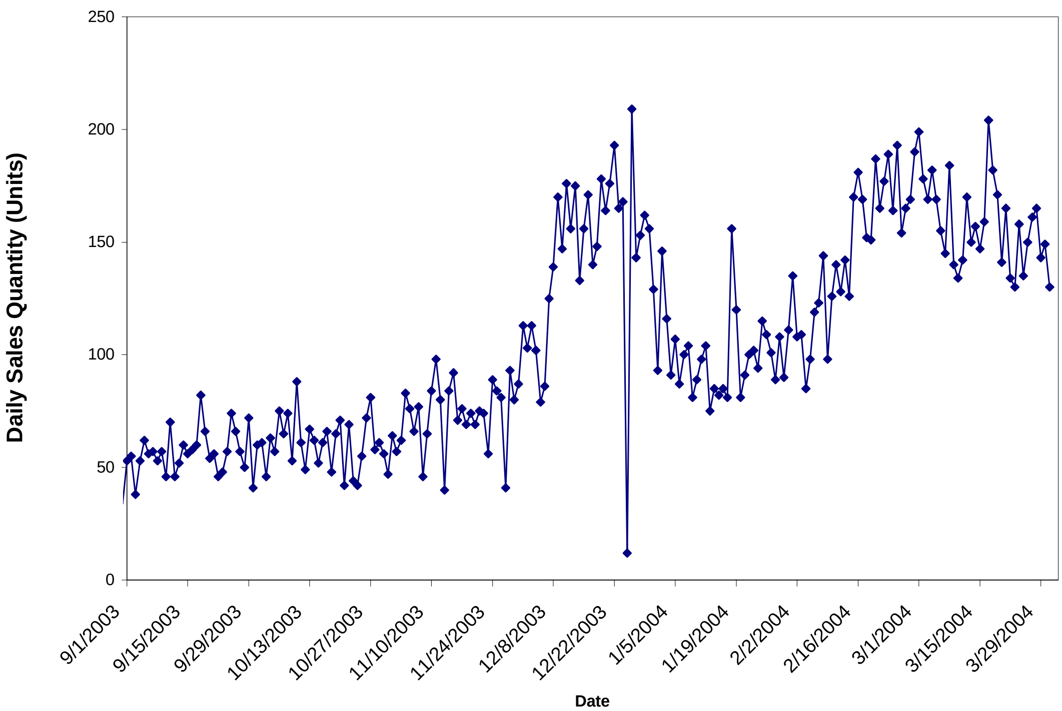

Figure 13 Electrolytes sales Allegheny County August 1, 2003 to March 31, 2004

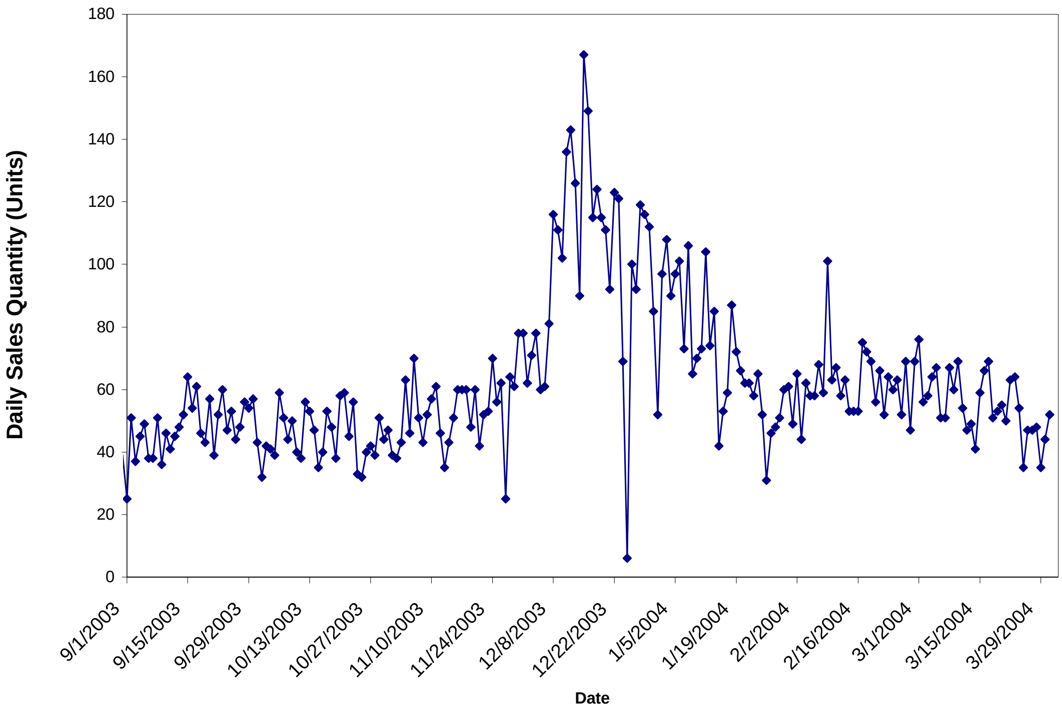

Figure 14 Thermometer sales Allegheny County August 1, 2003 to March 31, 2004

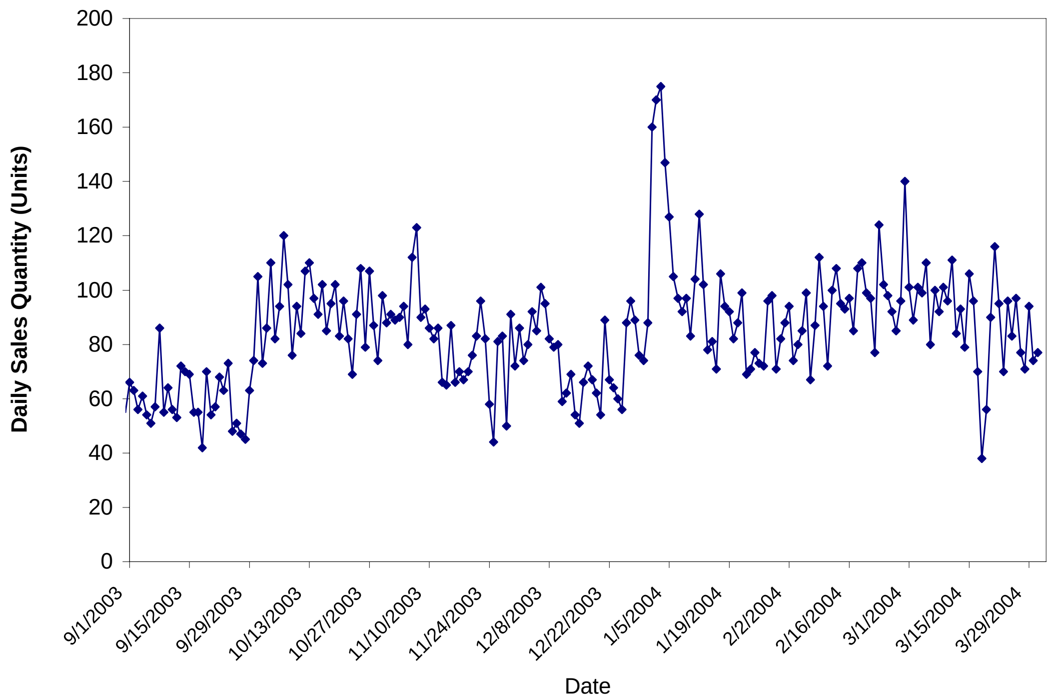

Figure 15 Constitutional Chief Complaints Allegheny County August 1, 2003 to March 31, 2004

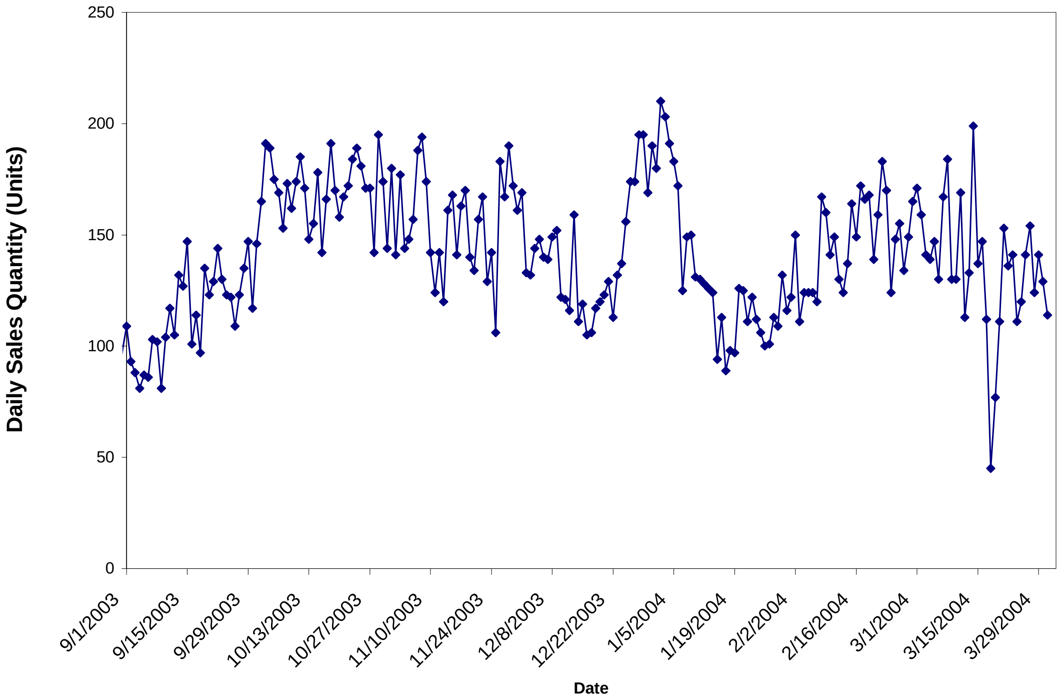

Figure 16 Respiratory Chief Complaints Allegheny County August 1, 2003 to March 31, 2004

Five prospective signals alarmed between December 8, 2003 and December 22, 2003 in Allegheny County. The first signal was produced by thermometer sales with a standard deviation of 4.29. Electrolyte sales followed with their first alarm on December 9, 2003 and anti-fever pediatric sales signaled for the first time on December 22, 2003. The alarm that had the highest standard deviation occurred in thermometer sales on December 15, 2003 with a standard deviation of 4.85. The prospective results are displayed in table 3.

Table 3 Prospective Alarms for Allegheny County, Pennsylvania

OTC	Date of signal	Location	Standard Deviations	# of signals for the 2 months prior to the outbreak	# of signals for the 2 months after the outbreak
Thermometers	12/8/2003	Allegheny	4.29	0	0
Electrolytes	12/9/2003	Allegheny	3.17	0	0
Thermometers	12/11/2003	Allegheny	4.17	0	0
Thermometers	12/15/2003	Allegheny	4.85	0	0
APP	12/22/2003	Allegheny	3	0	0

In the retrospective analysis the earliest signal in OTC sales for a false alarm rate of both two and four per year was in electrolyte sales, the signal was generated 26 days prior to widespread activity on October 27, 2003. Thermometers sales also signaled early for a false alarm rate of two per year. This signal was generated on November 4, 2003 in both ARIMA algorithms. Constitutional chief complaints signaled consistently on November 20, 2003 for a false alarm rate of two per year in each of the CuSum algorithms. The wavelet algorithm produced a signal on December 7, 2003 and both ARIMA algorithms signaled on November 30, 2003. Respiratory chief complaints signaled between November 28 and December 13, 2003 for a false alarm rate of two per year. This corresponds to a timeliness of six and 21 days. These results are displayed in table 4.

Table 4 Retrospective Alarms for Allegheny County, Pennsylvania

	CuSum								Wavelet		ARIMA			
	weight = 0.05 window = infinite		weight = 0.20 window = 10		weight = 0.20 window = infinite		weight = 0.05 window = 10		Wavelet		Adaptive		ARIMA (1,0,1)	
	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Anti-diarrheals	39	38	-15	-21	37	-19	112	108	**	**	**	34	**	34
APA	12	10	106	57	10	9	**	**	**	37	34	34	**	34
APP	11	11	**	30	30	9	30	21	**	16	**	9	**	9
Cough & Cold	-11	-18	**	**	-14	-14	**	**	-10	9	6	6	34	3
Electrolytes	-26	-26	11	10	**	-26	-15	-15	6	-33	9	-18	-11	-11
Thermometers	-16	-16	-16	-16	-16	-16	11	4	-12	-12	-18	-18	-18	-18
Constitutional	-2	-2	-2	-18	-2	-2	-2	-3	15	8	8	-5	8	8
Respiratory	17	14	21	18	18	15	20	15	6	6	6	6	6	6

4.3. Outbreak 3: Los Angeles County, California

The first laboratory confirmed case was identified in Los Angels County on October 10, 2003. On November 12, 2003 the County Health Department warned that the influenza season was beginning as a result of two additional positive cultures confirmed by their laboratory. On December 13, 2003 widespread activity was declared in California.

The time series curves for Los Angeles County, California are displayed for the same time period as the previous two outbreaks. September 1, 2003 to March 31, 2004. These time series data are in figures 17 through 22.

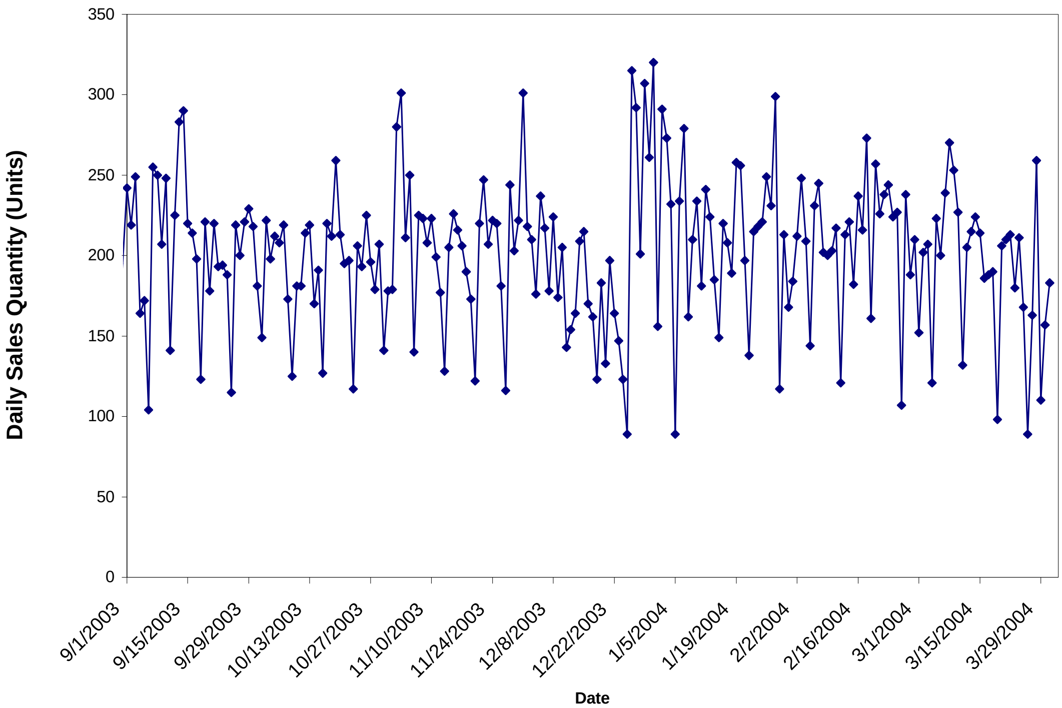

Figure 17 Anti-Diarrheal sales LA August 1, 2003 to March 31, 2004

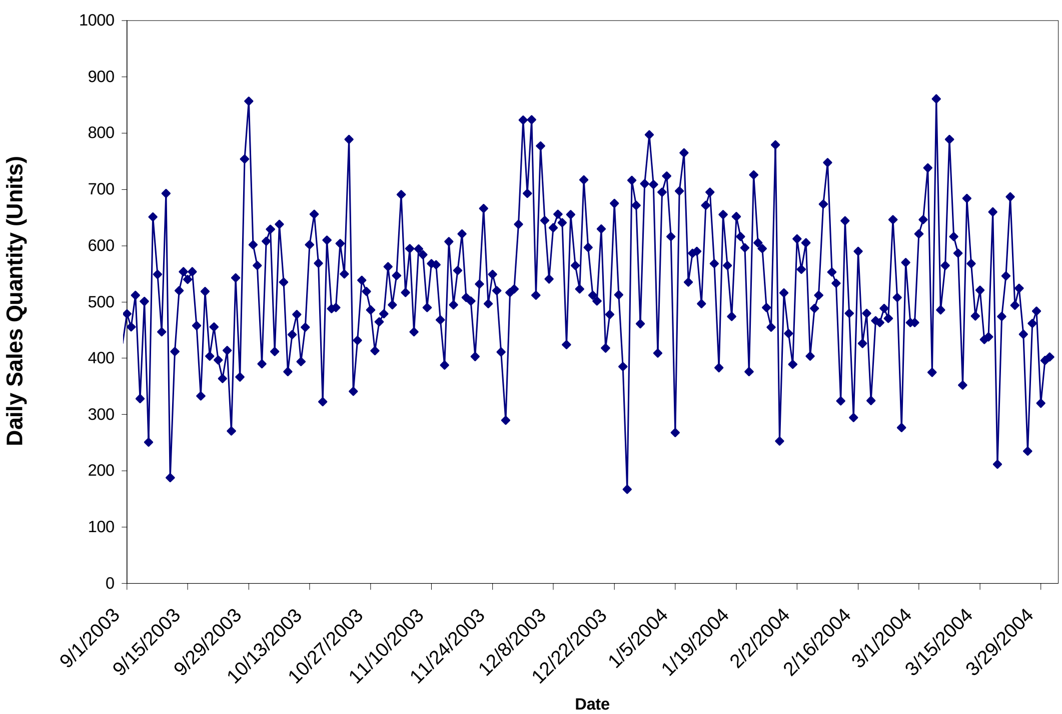

Figure 18 APA sales LA August 1, 2003 to March 31, 2004

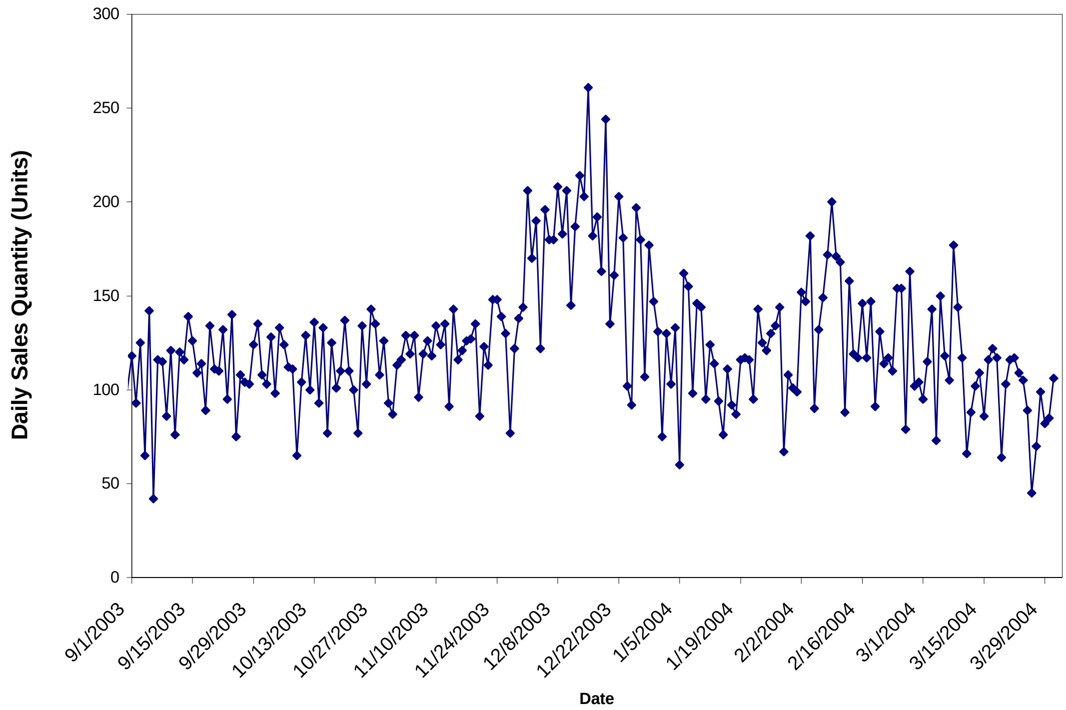

Figure 19 APP sales LA August 1, 2003 to March 31, 2004

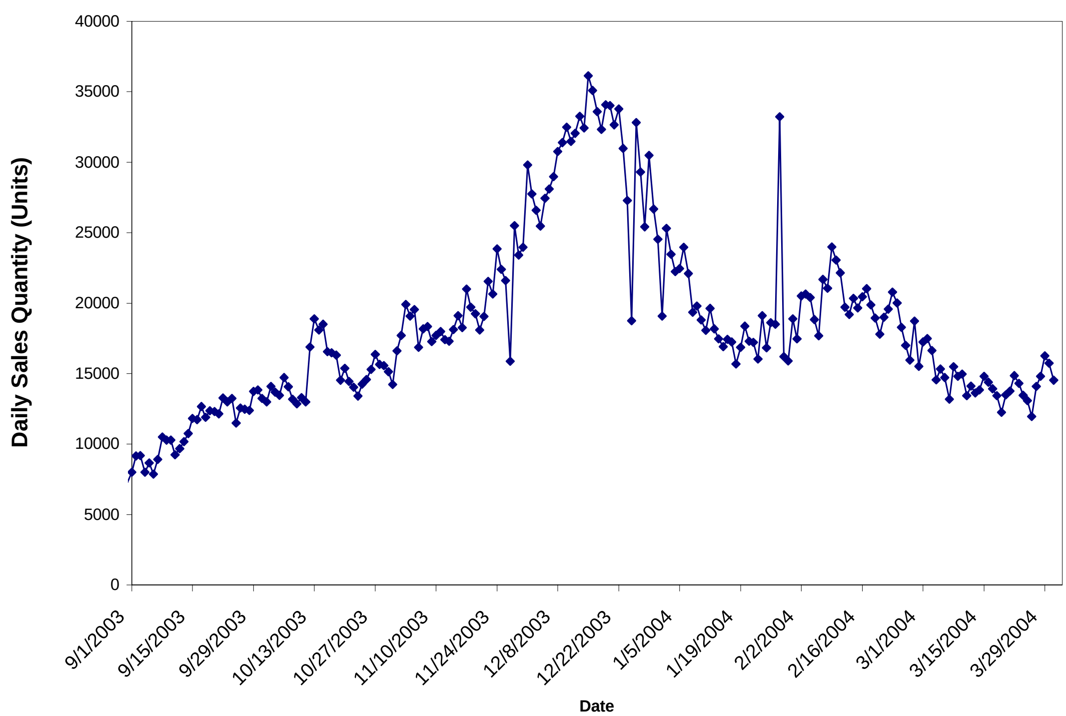

Figure 20 Cough/Cold sales LA August 1, 2003 to March 31, 2004

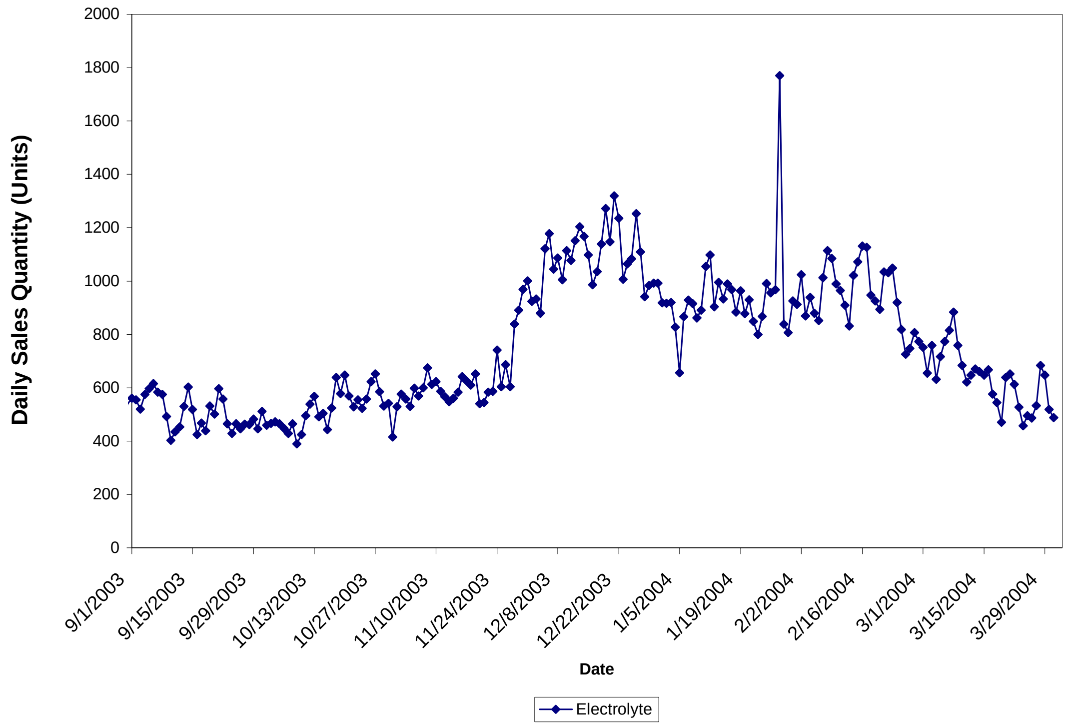

Figure 21 Electrolyte sales LA August 1, 2003 to March 31, 2004

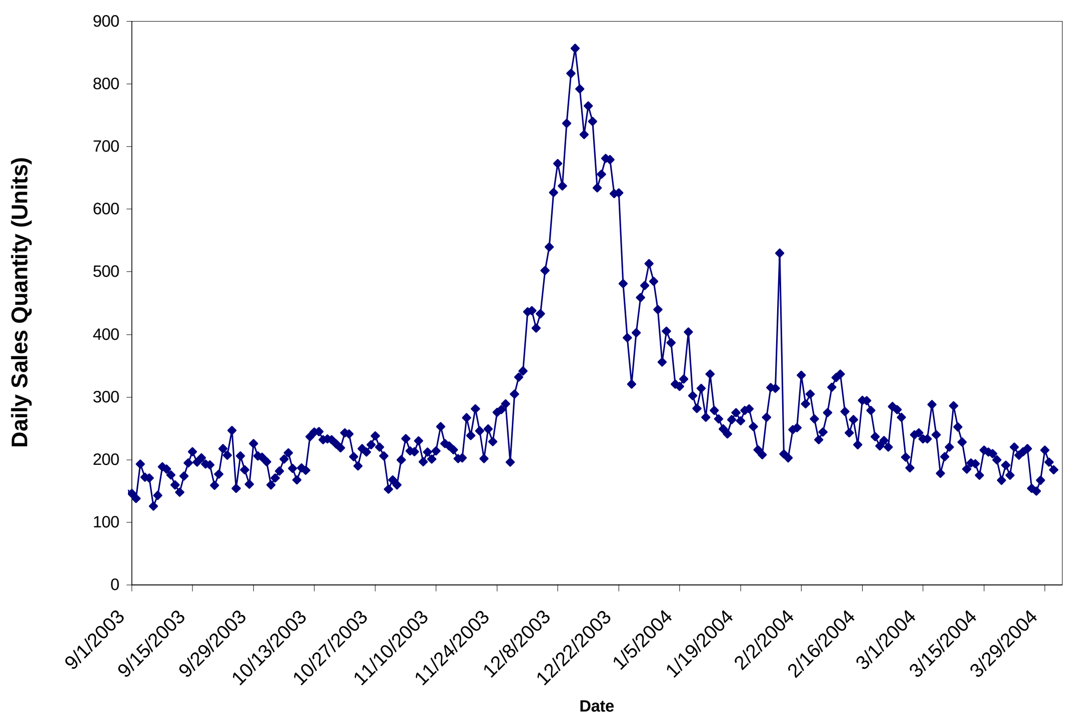

Figure 22 Thermometer sales LA August 1, 2003 to March 31, 2004

There were 11 prospective alarms in Los Angeles County. The first was on October 12, 2003 in the cough/cold category. Thermometer sales signaled on December 1, 2003 for the first time. None of the other categories signaled. The alarm with the highest standard deviation was on December 8, 2003 in thermometer sales with a standard deviation of 6.24. The results are displayed in table 5.

Table 5 Prospective Signals for Los Angeles County, California

OTC	Date of signal	Location	Standard Deviations	# of signals for the 2 months prior to the outbreak	# of signals for the 2 months after the outbreak
Cough/Cold	10/12/2003	Los Angeles	3.39	0	0
Cough/Cold	10/13/2003	Los Angeles	4.49	0	0
Cough/Cold	11/24/2003	Los Angeles	3.2	0	0
Thermometers	12/1/2003	Los Angeles	3.66	0	0
Cough/Cold	12/1/2003	Los Angeles	4.06	0	0
Thermometers	12/8/2003	Los Angeles	6.24	0	0
Cough/Cold	12/8/2003	Los Angeles	3.17	0	0
Thermometers	12/9/2003	Los Angeles	4.1	0	0
Thermometers	12/10/2003	Los Angeles	5.1	1	1
Thermometers	12/11/2003	Los Angeles	6.1	2	2
Thermometers	12/12/2003	Los Angeles	7.1	3	3

The earliest alarm for a false alarm rate of two per year was in electrolyte sales on October 18, 2003 in the ARIMA (1, 0, 1) algorithm, this is 56 days prior to widespread activity being declared. The cough and cold category signaled on November 3rd, 40 days prior to widespread activity being declared. Electrolytes and the cough and cold category also signal on November 4th 2003 for the same false alarm rate. The earliest signal for a false alarm rate of four per year was in electrolyte sales 56 days prior to widespread activity on October 18, 2003 in the ARIMA (1, 0, 1) algorithm. The retrospective analysis results are displayed in table 6.

Table 6 Retrospective Signals for Los Angeles County, California

	CuSum								Wavelet		ARIMA			
	weight = 0.05 window = infinite		weight = 0.20 window = 10		weight = 0.20 window = infinite		weight = 0.05 window = 10		Wavelet		Adaptive		ARIMA (1,0,1)	
	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Anti-diarrheals	**	**	-16	-16	14	-12	**	**	**	**	**	18	**	13
Anti-fever adult	**	-52	**	18	18	17	**	-52	**	**	13	-53	**	**
Anti-fever pediatric	-12	-18	-33	-33	-12	-12	-18	-29	**	**	-12	-12	**	-12
Cough & Cold	-39	-40	**	-39	-35	-35	-39	-39	**	-40	-15	-19	-15	-15
Electrolytes	-21	-27	-21	-27	-39	-39	-21	-37	-15	-19	-15	-15	-56	-56
Thermometers	-21	-22	-24	-55	-23	-23	-24	-24	-12	-13	-15	-26	-12	-12

4.4. Outbreak 4: Jefferson County, Kentucky

On the weekending of October 17, 2003 sentinel physicians in Jefferson County reported the first two cases of influenza. Widespread activity was declared on November 22, 2003. The peak of the outbreak occurred the week ending December 19, 2003 and the outbreak ended on March 12, 2003. The time series curves for Jefferson County, Kentucky are displayed in Figure 28 for the time period from September 1, 2003 to March 31, 2004.

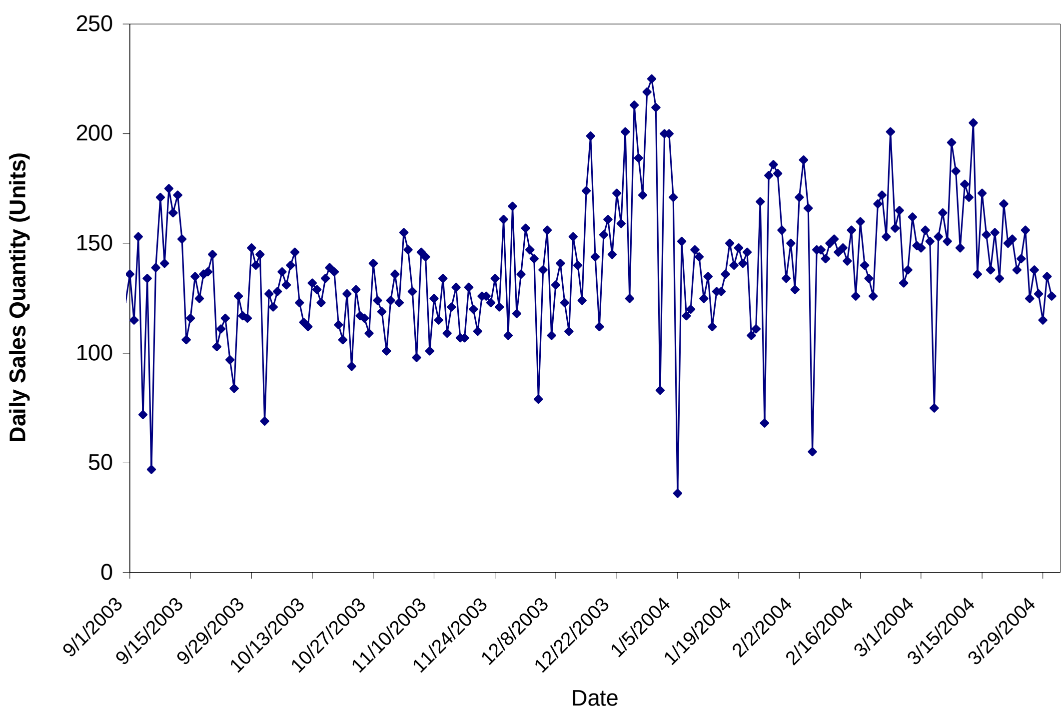

Figure 23 Anti-Diarrheal sales Jefferson County August 1, 2003 to March 31, 2004

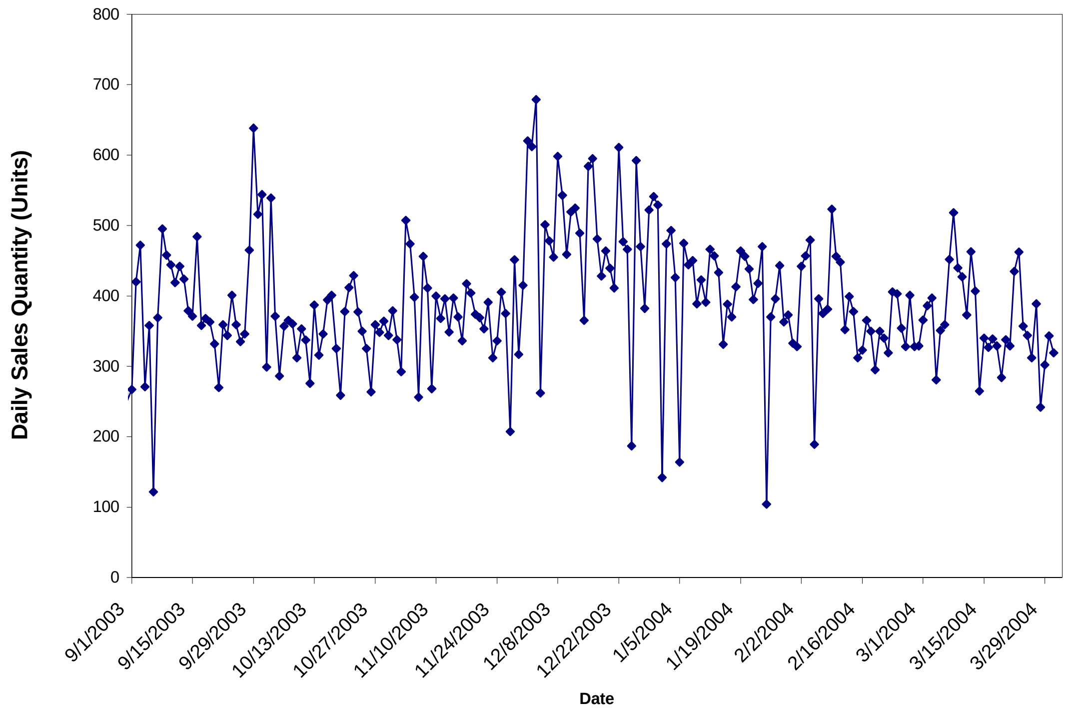

Figure 24 APA sales Jefferson County August 1, 2003 to March 31, 2004

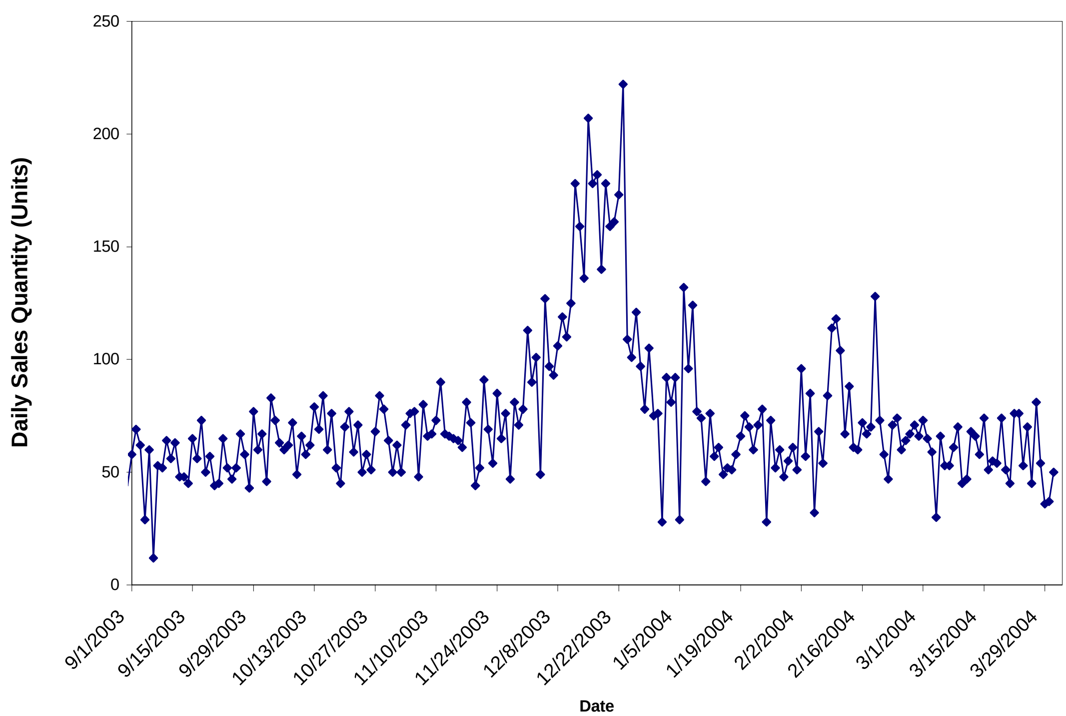

Figure 25 APP sales Jefferson County August 1, 2003 to March 31, 2004

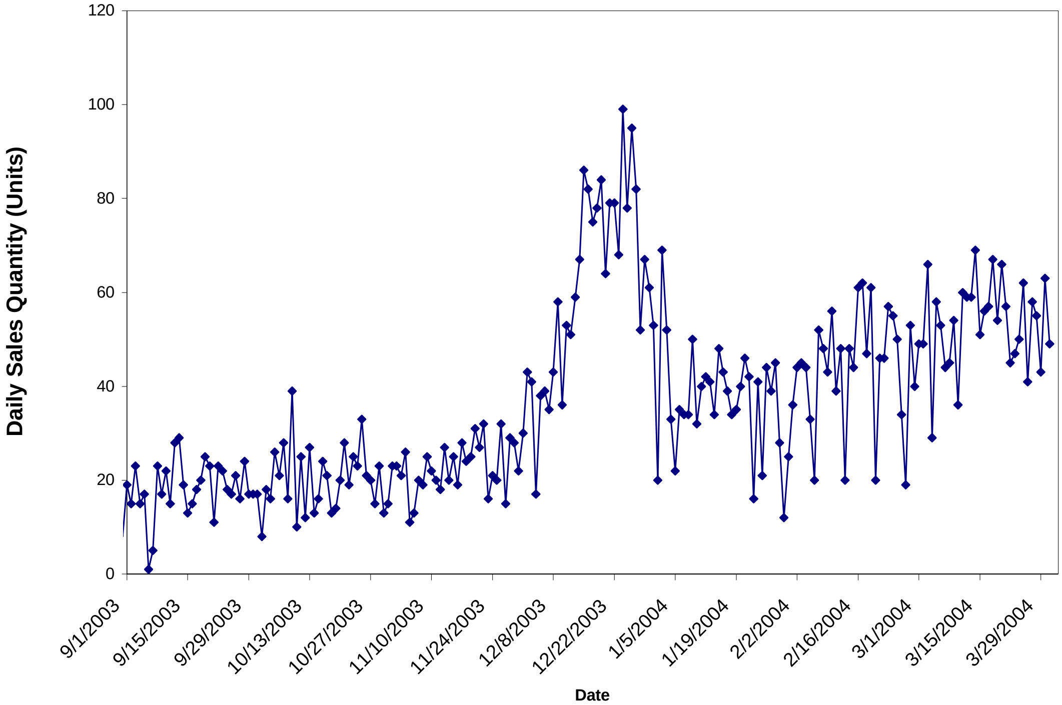

Figure 26 Cough/Cold sales Jefferson County August 1, 2003 to March 31, 2004

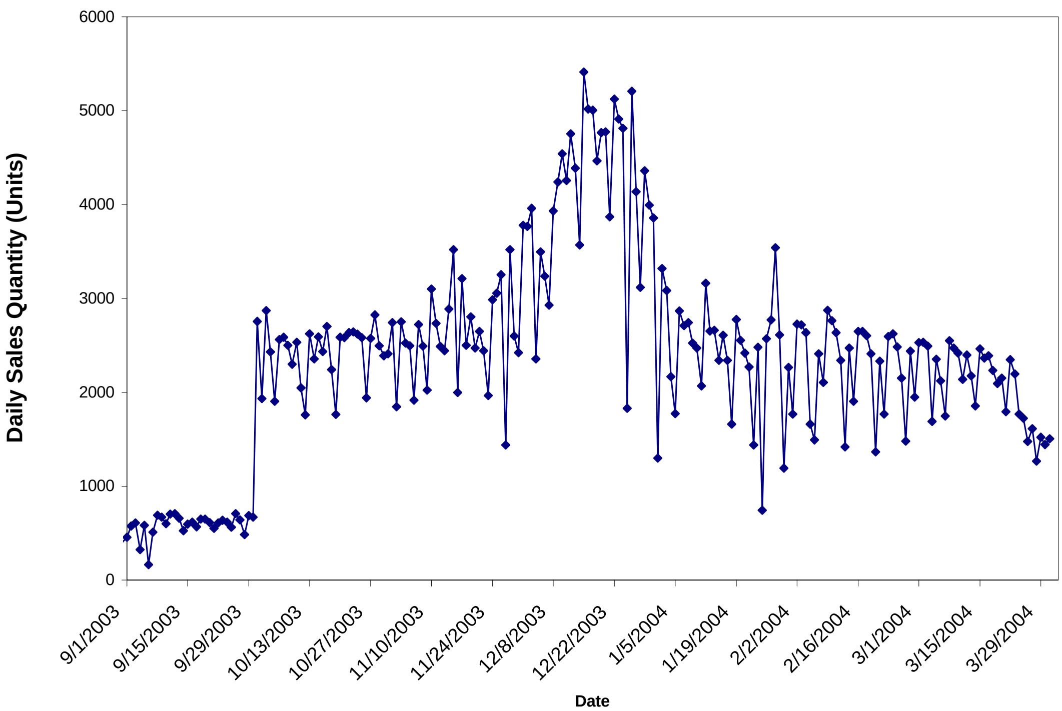

Figure 27 Electrolytes sales Jefferson County August 1, 2003 to March 31, 2004

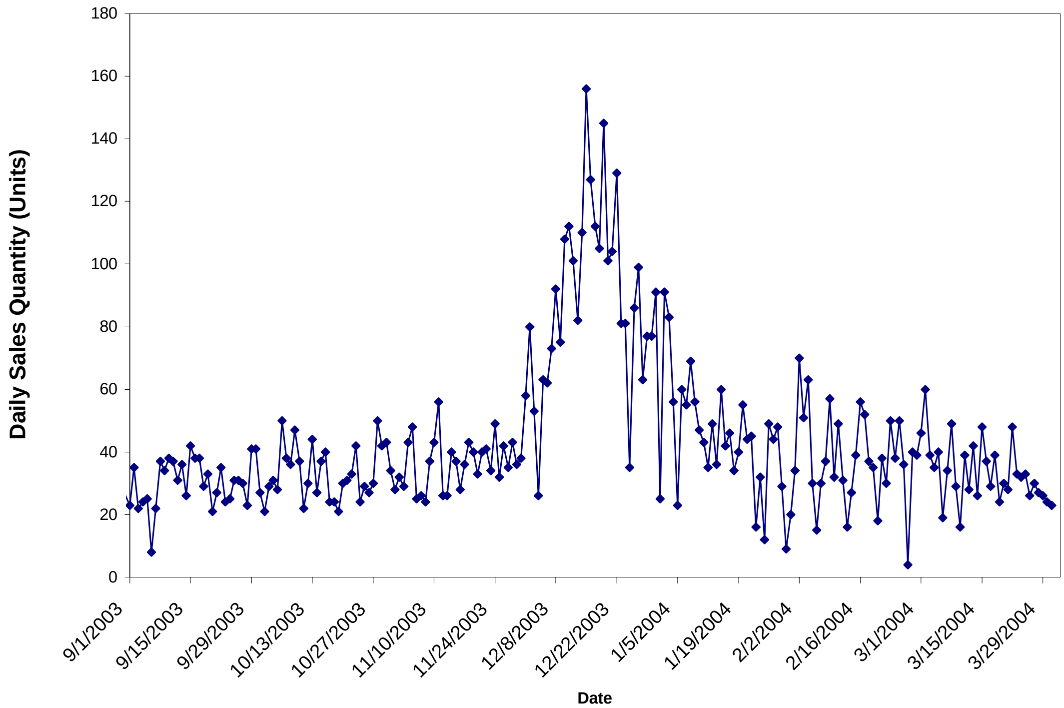

Figure 28 Thermometer sales Jefferson County August 1, 2003 to March 31, 2004

The prospective analysis produced eight alarms. Each of these alarms signaled between December 7, 2003 and December 15, 2003. The first alarm to signal was in thermometer sales. In addition to thermometer sales electrolytes and the cough/cold category first signaled on December 9, 2003 and December 15, 2003, respectively. The anti-diarrheal and anti-fever pediatric categories did not alarm prospectively. The alarm that had the highest number of standard deviations occurred on December 15, 2003 in thermometers sales with 6.46 standard deviations. The results are displayed in table 7.

Table 7 Prospective Signals for Jefferson County, Kentucky

OTC	Date of signal	Location	Standard Deviations	# of signals for the 2 months prior to the outbreak	# of signals for the 2 months after the outbreak
Thermometers	12/7/2003	Jefferson	3.38	0	0
Thermometers	12/8/2003	Jefferson	4.34	0	0
Electrolytes	12/9/2003	Jefferson	3.23	1	0
Thermometers	12/10/2003	Jefferson	4.82	0	0
Thermometers	12/11/2003	Jefferson	3.23	0	0
Thermometers	12/15/2003	Jefferson	6.46	0	0
Cough/Cold	12/15/2003	Jefferson	3.68	0	0
Electrolytes	12/15/2003	Jefferson	5.12	1	0

The earliest alarm generated was in anti-fever pediatric sales on October 15, 2003. This was 38 days prior to widespread activity in the CuSum algorithm with a weight of 0.05 and a window of 10. The earliest alarm for a false alarm rate of four per year was also for anti-fever pediatric sales seen in the CuSum algorithm using a weight of both 0.05 and 0.20 with a window of 10. The retrospective analysis results are displayed in table 8.

Table 8 Retrospective Signals for Jefferson County, Kentucky

	CuSum								Wavelet		ARIMA			
	weight = 0.05 window = infinite		weight = 0.20 window = 10		weight = 0.20 window = infinite		weight = 0.05 window = 10		Wavelet		Adaptive		ARIMA (1,0,1)	
	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Anti-diarrheals	34	32	-19	-19	38	24	35	32	**	**	**	34	**	34
Anti-fever adult	10	10	-19	-19	10	10	11	11	**	**	34	34	**	34
Anti-fever pediatric	14	11	20	-38	19	17	-38	-38	**	**	9	9	**	9
Cough & Cold	8	7	8	6	6	-33	8	6	**	**	47	6	-11	-11
Electrolytes	0	-28	0	0	11	10	-2	-28	**	**	6	6	34	3
Thermometers	10	10	-23	-22	7	-24	9	-23	10	10	-18	-18	-18	-18

.

The AMOC curves were generated for four data streams including APP, electrolytes, cough and cold, and constitutional chief complaints. These OTC data streams had the earliest OTC detection in at least one of the four influenza outbreaks. The constitutional chief complaints data stream was chosen since it had the earliest detection in the Pittsburgh and Salt Lake outbreaks for ED chief complaints. Figures 5.1 through 8.3 display the AMOC curves and will be further discussed in the next section.

5. Discussion

A visual inspection of the times series graphs reveal the effect that influenza has on OTC sales and ED chief complaints. What cannot be determined visually in real time is the point in time at which this data actually becomes anomalous. What is consistent with the time series is that a gradual increase is seen as opposed to a one day spike.

There are limitations to this experiment. First, only six months of historical data are available to set the false alarm rates. Second, the market share in each of these counties is not the same and the data collection method does not take this into account. Lastly, other factors that influence OTC sales cannot be extracted, so if another outbreak occurred during this time it may appear as if the sales are the result of the influenza season when it actually is some other phenomenon. With that said, the conclusions made from this study are preliminary and the capability to assess the exact time of the start of the influenza season will require some additional inputs, mainly a longer time frame of historical data, one that spans years not months.

Another aspect to be considered is the determination of the optimal false alarm rate to use for this analysis. The timeliness is increased by many days when the false alarm rate is increased. A primary example of this can be seen with the electrolytes data stream in Los Angeles County and the adaptive ARIMA algorithm in figure 29.

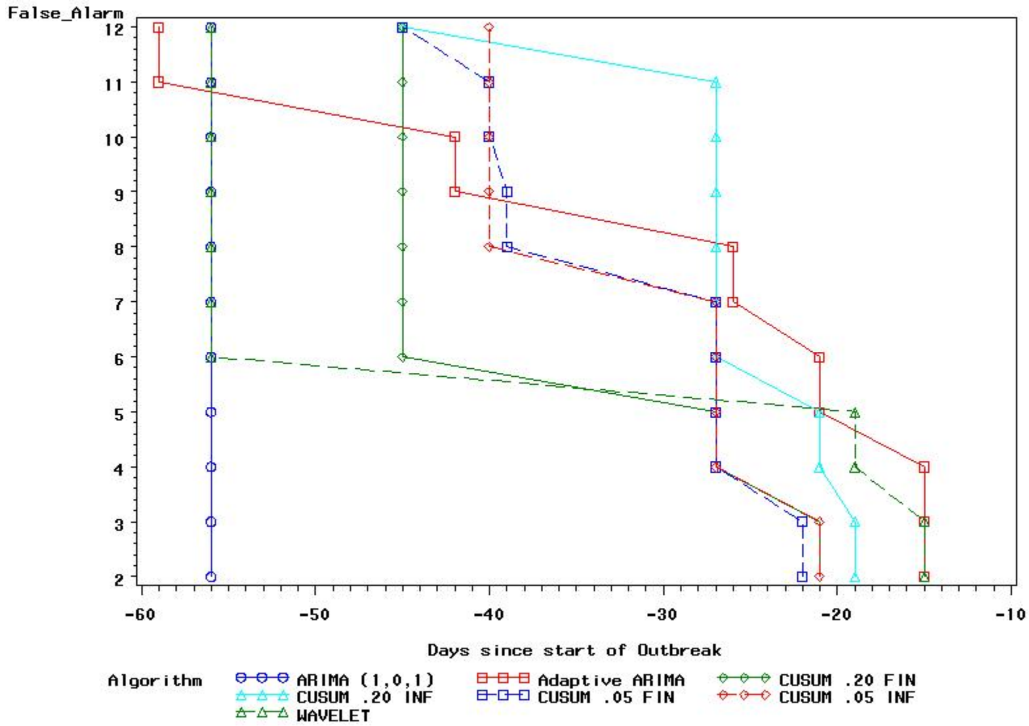

Figure 29 AMOC Electrolytes Los Angeles County, California

For a false alarm rate of 12 per year a signal is generated on October 15th, 59 days before the reference date, but for a false alarm rate of two per year the signal is not generated until November 28th, only 15 days before the reference date. The difference in detection is 44 days earlier for a false alarm rate of 12 compared to a false alarm rate of two. Twelve false alarm rates per year is equal to one alarm per month, and if public health officials are monitoring these data the alarms each month will most likely be an expectation as opposed to an alarm. A false alarm rate as low as two per year reduces the timeliness and the reaction time that public health has to respond to the outbreak. Similar situations can be seen in various other AMOC. These are displayed in figures 30 through 42.

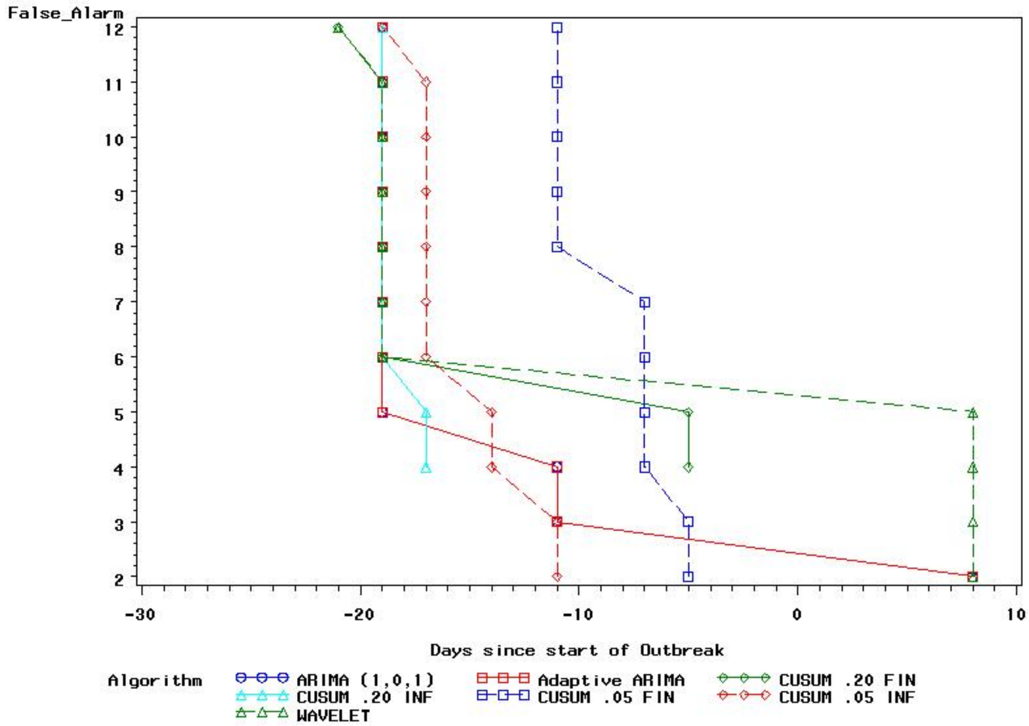

Figure 30 AMOC APP Salt Lake County, Utah

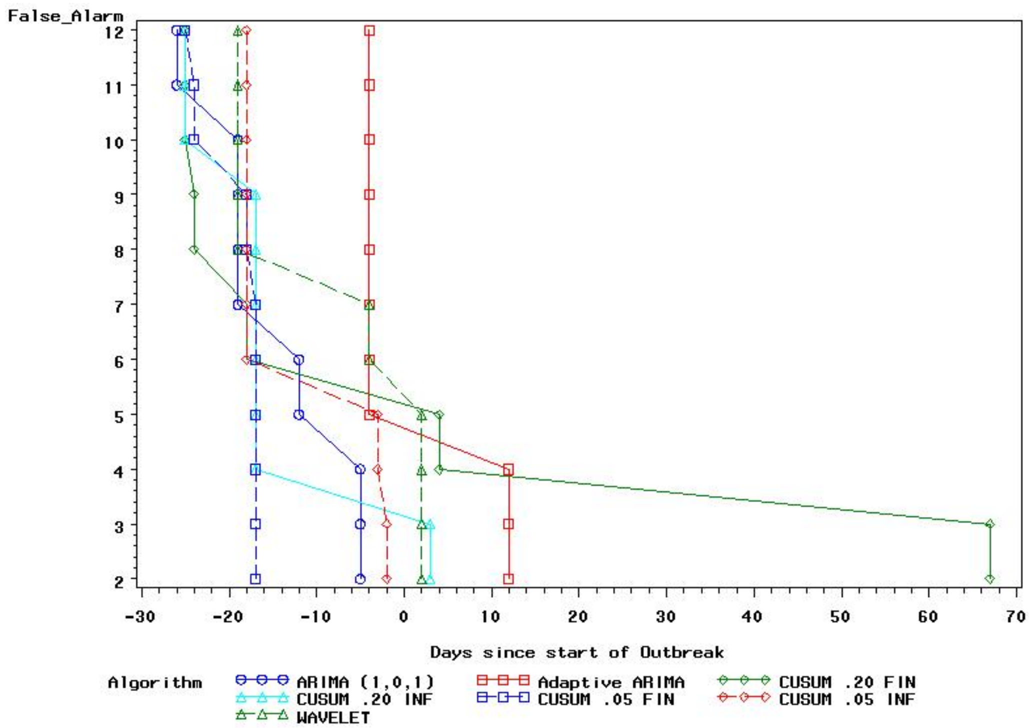

Figure 31 Cough & Cold Salt Lake County, Utah

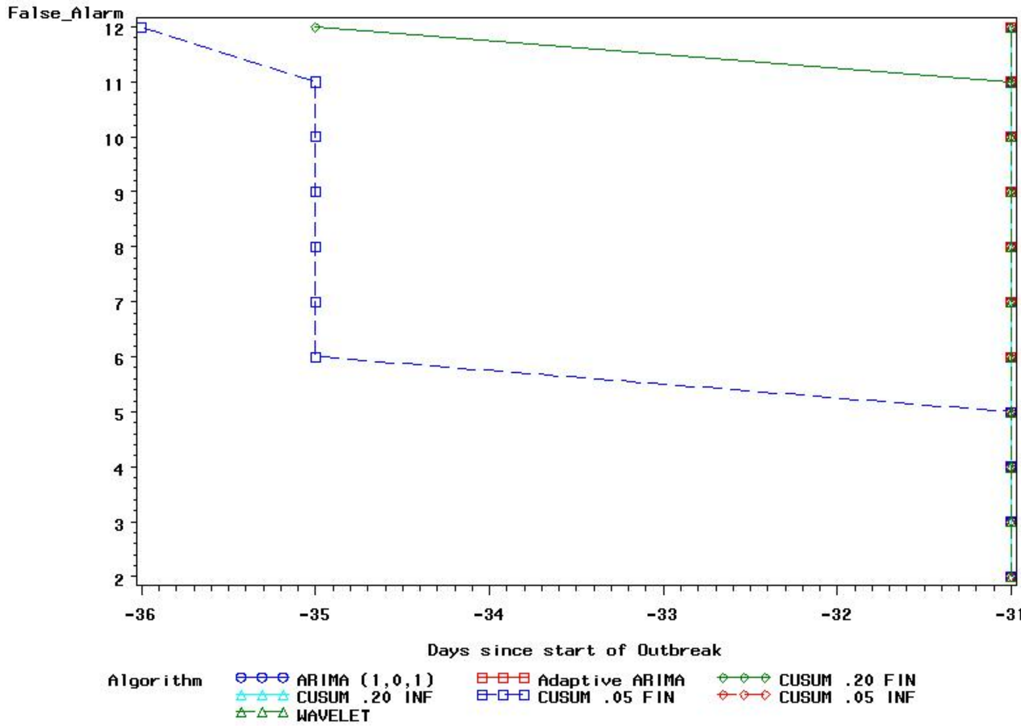

Figure 32 AMOC Electrolytes Salt Lake County, Utah

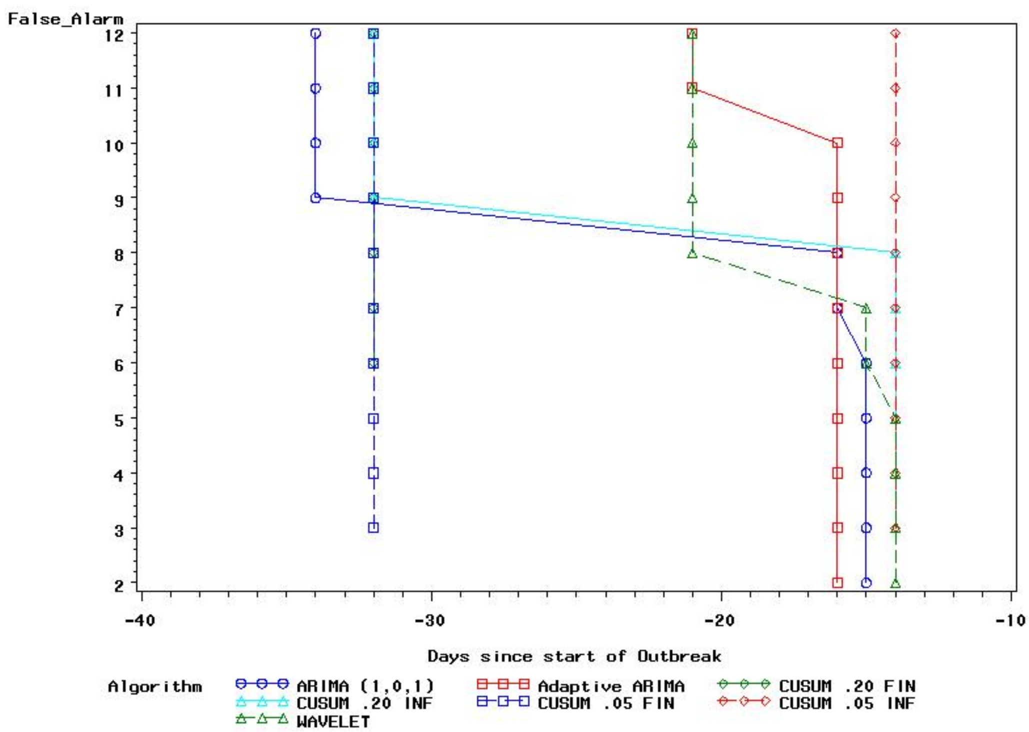

Figure 33 AMOC Constitutional Chief Complaints Salt Lake County, Utah

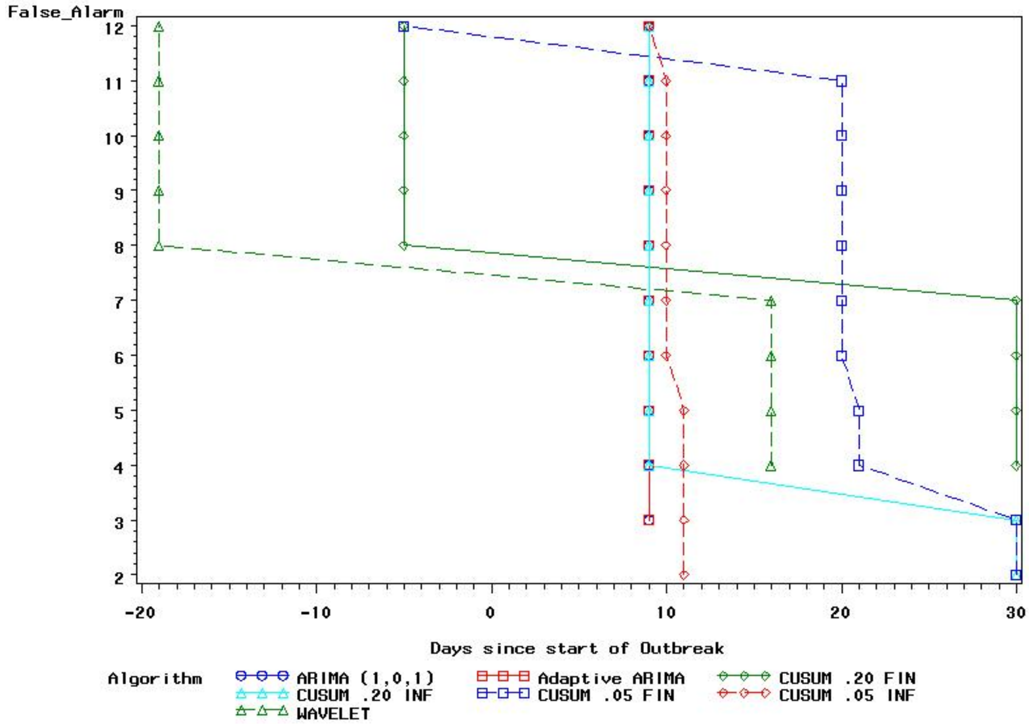

Figure 34 AMOC APP Allegheny County, Pennsylvania

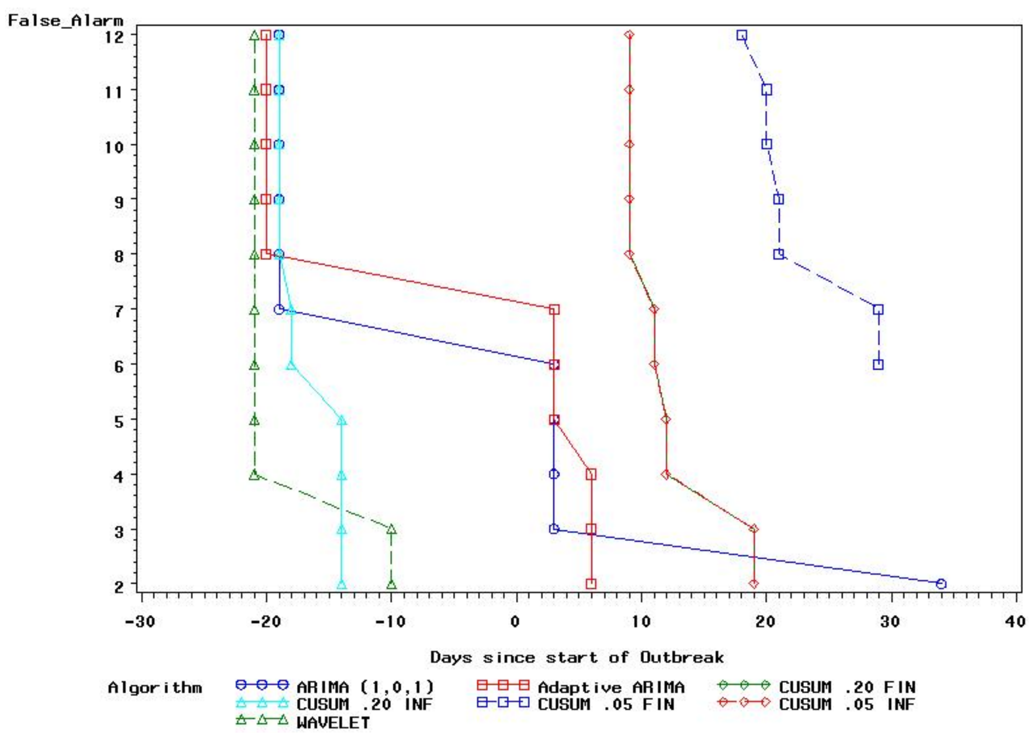

Figure 35 AMOC Cough & Cold Allegheny County, Pennsylvania

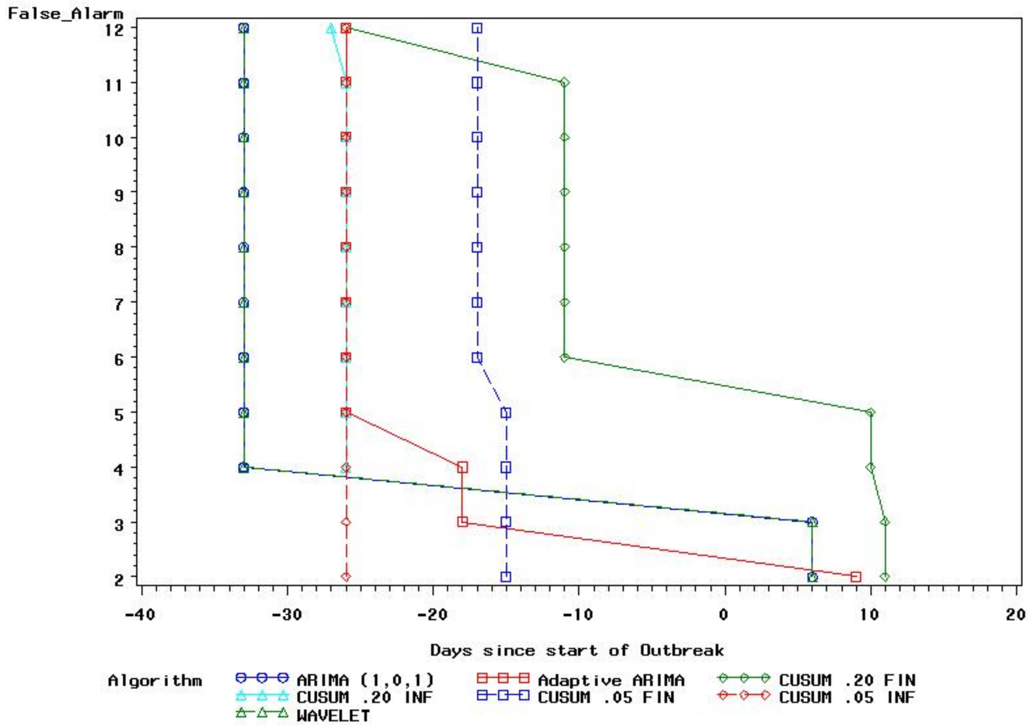

Figure 36 AMOC Electrolytes Allegheny County, Pennsylvania

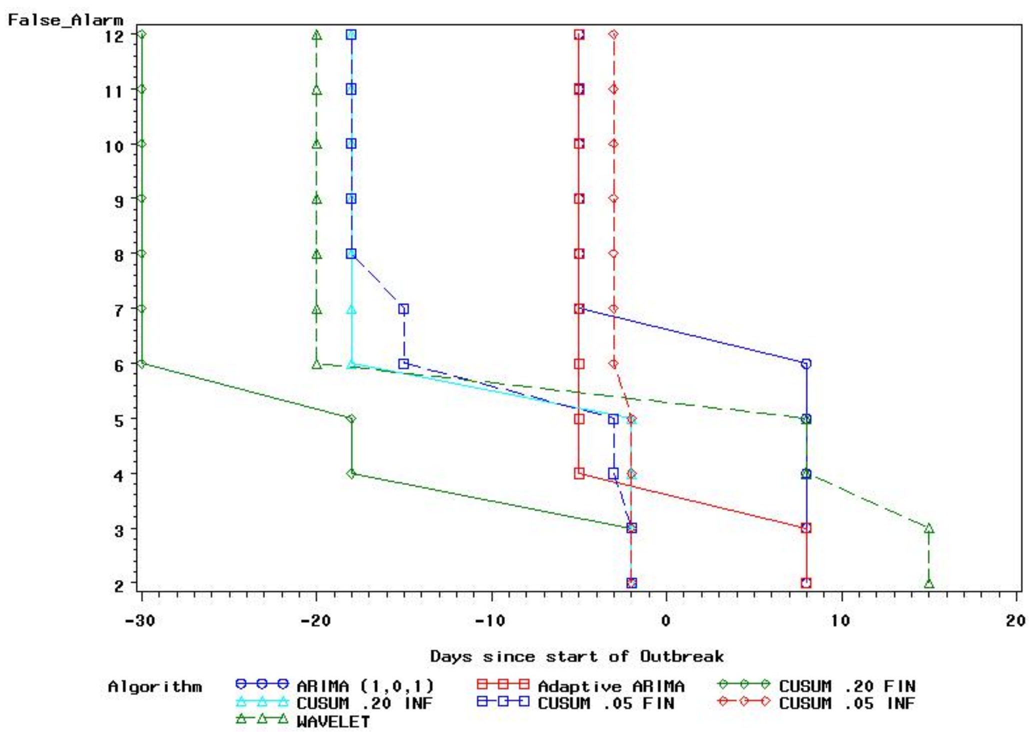

Figure 37 AMOC Constitutional Chief Complaints Allegheny County, Pennsylvania

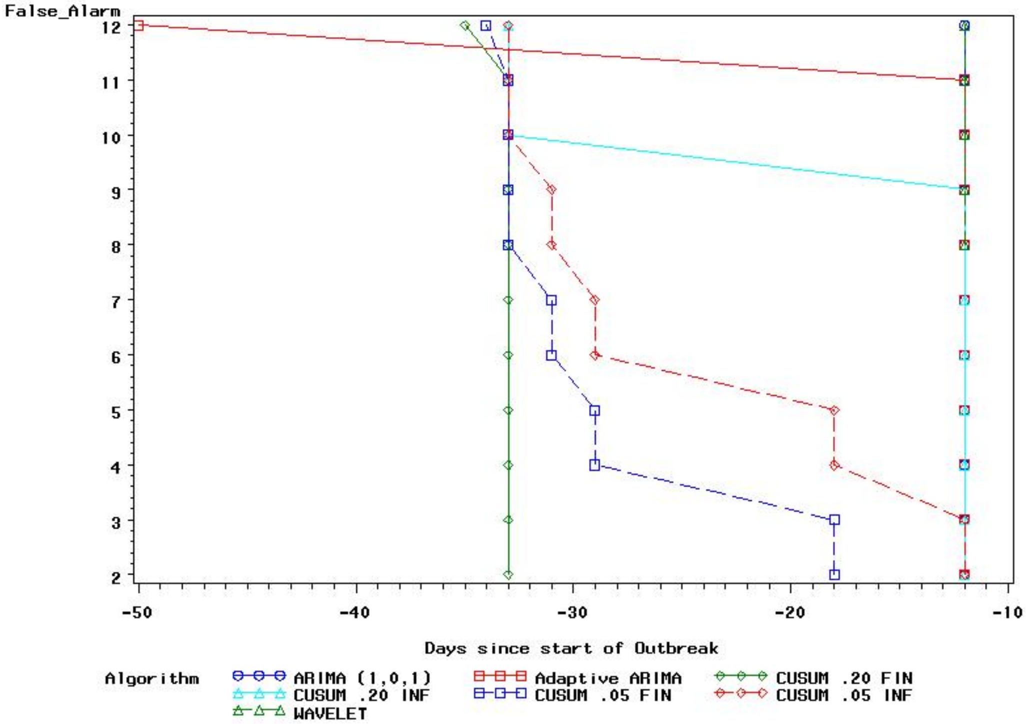

Figure 38 AMOC APP Los Angeles County, California

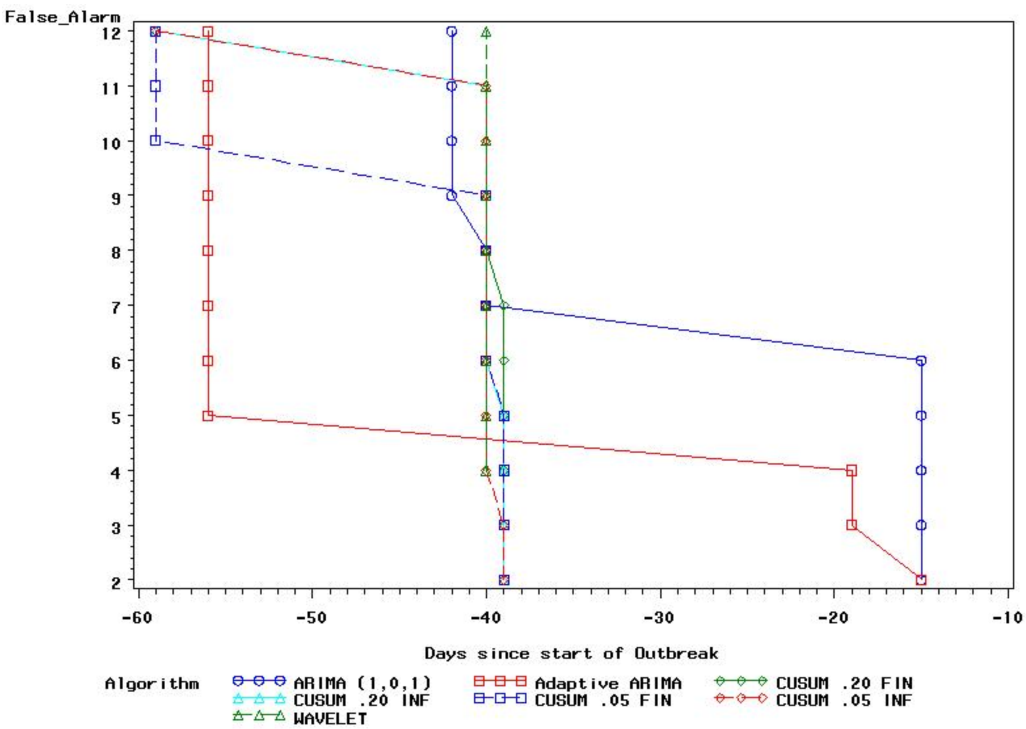

Figure 39 AMOC Cough and Cold Los Angeles County, California

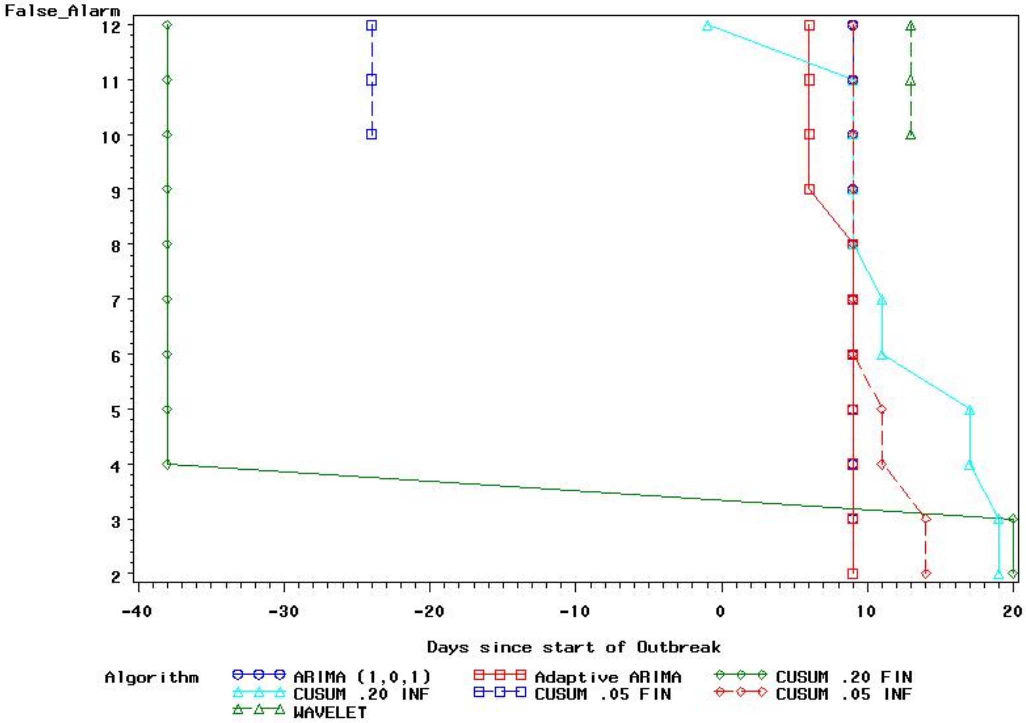

Figure 40 AMOC APP Jefferson County, Kentucky

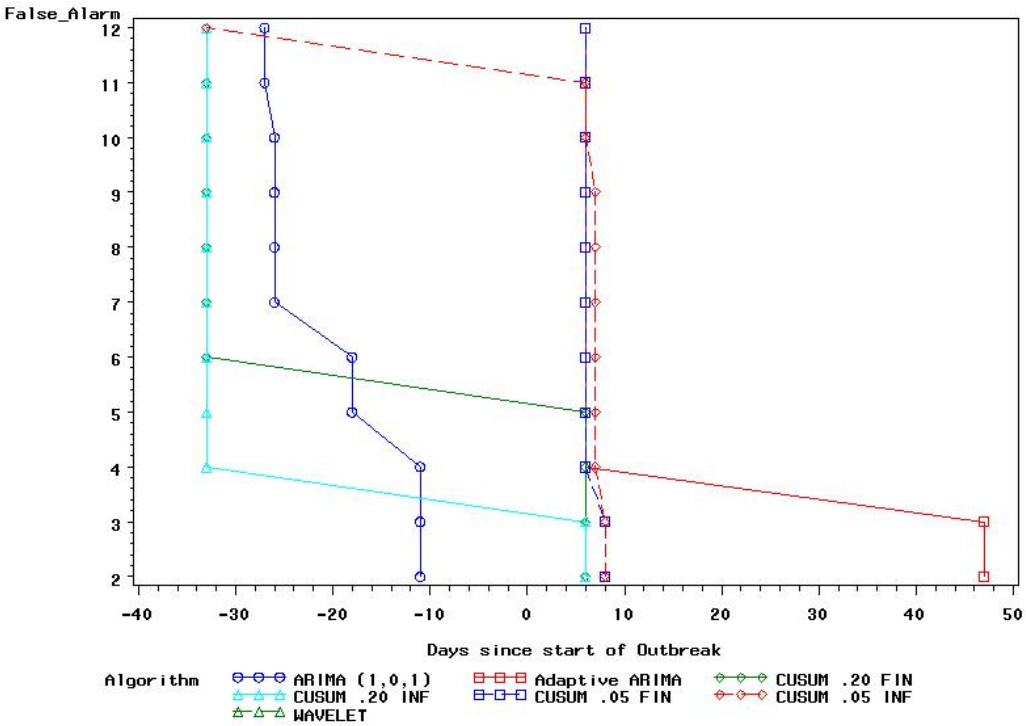

Figure 41 AMOC Cough and Cold Jefferson County, Kentucky

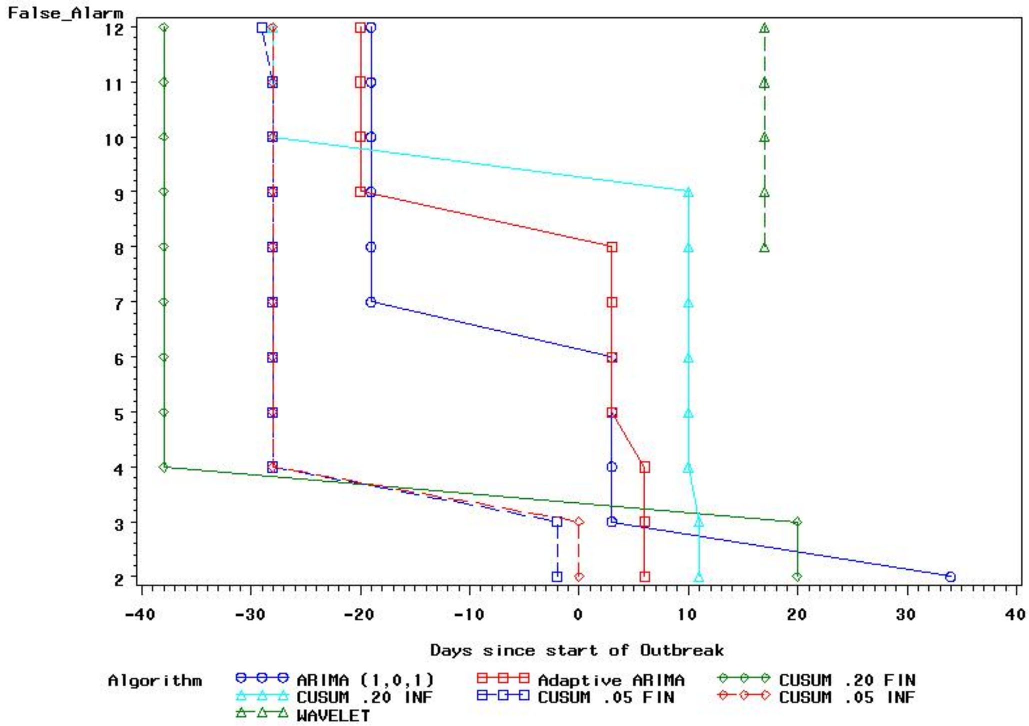

Figure 42 AMOC Electrolytes Jefferson County, Kentucky

The signals produced by false alarm rates of four and two per year are emphasized since this appears to be a reasonable amount of yearly alarms. False alarm rates of three per year are also reasonable. But, since there is only six months of historical data, the timeliness of this false alarm rate is equal to either the timeliness of the false alarm rate of two or four per year.

It is difficult to make conclusions regarding which data stream is best to use for monitoring. As stated before these outbreaks are not identical, have noisy data and only one influenza outbreak is investigated. A visual inspection of the time series curves in figures 1 through 28 shows some consistency among the data streams. The anti-diarrheal and APA data streams do not show an explicit increase during the outbreak like the other OTC data streams do. These two data streams are variable and with sales continually going up and down during the outbreak. The cough/cold category does show a gradual increase in sales during the outbreak, but this increase can begin much earlier then the influenza outbreak. This phenomenon is apparent especially in the LA data. The increase begins in early September. This may be due to asthmatic and allergy conditions seen in the late summer and early fall. This data stream may give several signals before the start of the influenza season due to these factors. Thermometers, electrolytes, APP, and the two chief complaints investigated show a gradual increasing trend during the start of the influenza season and are the primary data streams of interest and are recommended to be monitored by public health.

Electrolyte sales were the first data stream to signal in three of the four outbreaks for a false alarm rate of two per year. This data stream appears to be the one best suited for influenza monitoring since it consistently alarmed first in three of the four outbreaks. Jefferson County was the exception. The earliest data stream to signal in that outbreak was APP. This data stream signaled 38 days prior to widespread activity in only one algorithm (CuSUM weight = 0.05, window = 10). The rest of the alarms for a false alarm rate of two per year had a timeliness between nine and 14 days after widespread activity was declared.

6. Conclusions

Influenza outbreaks can be detected monitoring OTC and hospital chief complaint data at an earlier time then in traditional methods Based on the information presented and the previous analyses, electrolyte sales appear to be the most sensitive measure of an over the counter preparation that is consumed at a greater rate during an influenza outbreak A determination of which algorithm is best suited to analyze the data is a more difficult task. Across each outbreak there is not one algorithm that consistently performs better then the others. The only generalization that can be made is that some data streams perform better then others. The data stream's signaling performance cannot be strictly tied to one algorithm. These preliminary results show that the data streams ability to signal is independent of the algorithm and dependent on the data stream.

6.1. Public Health significance of this work:

In the field of Public Health, lead time is very important as it permits professionals to be best prepared to respond to an increased need for hospital utilization and use of medical resources, Recent disasters in Louisiana and Thailand underscore the importance of preparedness and the ability to have resources available to deal with public health issues related to such crises in a timely manner. Syndromic surveillance and monitoring of over the counter sales of routine preparations are an important tool in the arsenal of skills and techniques that the Public Health establishment can count on to minimize severity of illness and to encourage primary, secondary and tertiary preventive strategies and to act in a timely manner to alert the public.

Influenza is only one potential health threat that has the ability to be detected. As research in the field of syndromic surveillance increases other disease outbreak's investigations will be able to determine which data streams and algorithms are best for the detection of a particular outbreak.

APPENDIX A

Description of Algorithms

ARIMA

The ARIMA model is a univariate autoregressive moving average model that predicts values from time series data and is a linear combination of its own past values and errors. This model also uses equally spaced intervals for forecasting, specifically days.

CuSum

The CuSum algorithm utilizes exponential weighted averages to detect shifts from the mean. This algorithm also uses historical data to predict a forecast. The model was adapted to detect one day forecasts and utilizes the forecasts residuals to determine if the data is anomalous.

WAVELET

The wavelet transform takes a time-series and transforms it into several different frequency bands. After the time-series is placed into several resolutions a one-step independent prediction is made for each resolution. The sum of each prediction at each resolution is summed to obtain the expected value for the current day.

BIBLIOGRAPHY

1. Yuan CM, Love S, Wilson M Syndromic Surveillanceat Hospital Emergency Departments – Southeastern Virginia, September 2004 MMWR 53(Suppl) 56-58

2. Metzger K, Hajat A, Crawford M, Mostashari F How Many Illnesses Does One Emergency Department Visti Represent? Using a Population Based Telephone Survey to Estimate the Syndromic Multiplier, September 2004 MMWR 53(Suppl) 106-111

3. Snacken R Weekly Monitoring of influenza in belgium (1993-1995) Pharmacoeconomics 1996; 9 Suppl 3:34-7

4. Preventing emerging infectious diseases: a strategy for the 21st century. Atlanta, GA: US Department of Health and Human Services 1998.

5. CDC. Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. MMWR 2001; 50:889-93.

6. Baxter R, Rubin R, Steinberg C, Carroll C, Shapiro J, Yang A Assessing core capacity for infectious disease surveillance. Final Report Prepared for: Office of the Assistant Secretary for Planning and Evaluation, DHHS, The Lewin Group, Inc 2000

7. Lazarus R, Kleinman K, Dashevsky I, Demaria A, Platt R Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection, October 2001 BMC Public Health

8. http://www.cdc.gov/nedss

9. M'ikanatha N, Southwell B, Lautenbach E, Automated Laboratory Reporting of Infectious Diseases in a Climate of Bioterrorism, September 2003, Emerging Infectious Disease

10. Baxter R, Rubin R, Steinberg C, Carroll C, Shapiro J, Yang A. Assessing core capacity for infectious disease surveillance. 2000; The Lewin Group, Inc.

11. Centers for Disease Control and Prevention. Notice to readers: ongoing investigation of anthrax—Florida, October 2001. MMWR 2001;50:877

12. Mandl K, Overhage M, Wagner M, Lober WB, Sebastiani P, Mostashari F, Pavlin JA, Gesteland P, Treadwell T, Koski E, Hutwagner L, Buckeridge DL Aller RD, Grannis S Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience; J Am Med Inform Assoc. Mar-Apr 2004 Journal of American Medical Informatics 11(2): 141–150.

13. Miller b, Kassenborg H, Dunsmuir W, Griffith J, Hadidi M, Nordin Jm Danila R Syndromic Surveillance for Influenzalike Illness in an Ambulatory Care Network, October 2004: Emerging Infectious Diseases Vol. 10, No. 10

14. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. Ann Emerg Med. 2003;41:8:447–52.

15. Lober WB, Trigg LJ, Karras BT, Bliss D, Ciliberti J, Duchin JS, et al. Syndromic surveillance using automated collection of computerized discharge diagnosis. J Urban Health. 2003;80(2 Suppl 1):i97–106.

16. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for communitywide influenza-like illness, New York City. J Urban Health. 2003;80(2 Suppl 1):i43–9.

17. Greenko J, Mostashari F, Fine A, Layton M. Clinical evaluation of the Emergency Medical Services (EMS) Ambulance Dispatch-Based Syndromic Surveillance System, New York City. J Urban Health. 2003;80(2 Suppl 1):i50–6.

18. Platt R, Bocchino C, Caldwell B, Harmon R, Kleinman K, Ritzwoller DP, et al. 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):i25–31.

19. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Pavlin J, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENSE II). J Urban Health. 2003;80(2 Suppl 1):i32–42.

20. Goldenberg A, Shmueli G, Caruana R, Fienberg S. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci U S A. 2002;99:5237–40

21. Rodman J, Frost F, Jakubowski W. Using nurse hot line calls for disease surveillance. Emerg Infect Dis. 1998;4:329–32.

22. Henning Kelly J What is Syndromic Surveillance?, MMWR September 2003: 53:7-11.

23. Foldy SL, Linking Better Surveillance to Better Outcomes, MMWR September 2003:12-17.

24. Jorm et al. “Watching the Games: public Health surveillance for the Sydney 2000 Olympic games.” J Epidemiology Community Health , 2003, 57:102-108.

25. Merchant GL, Mower WR, Talan DA. Influenza: ED considerations for the 1997–98 season. Ann Emerg Med 1997;30:692-694

26. Neuzil KM, Reed GW, Mitchel EF, et al. Influenza-associated morbidity and mortality in young and middle-aged women. JAMA 1999;281:901-907.

27. Barenfanger J, Drake C, Leon N, Mueller T, Troutt T. Clinical and financial benefits of rapid detection of respiratory viruses: an outcomes study. Journal of Clinical Microbiology August 2000; 38(8):2824-8.

28. Woo P C, Chiu S S, Seto W, Peiris M. Cost-effectiveness of rapid diagnosis of viral respiratory tract infection in pediatric patients Journal of Clinical Microbiology 1997;35:1579–1581.

29. Kohane IS. The contributions of biomedical informatics to the fight against bioterrorism. *J Am Med Inform Assoc.* 2002;9:116–9.

30. Teich JM, Wagner MM, Mackenzie CF, Schafer KO. The informatics response in disaster, terrorism, and war [comment]. 2002:97–104.

31. Koplan J. CDC's strategic plan for bioterrorism preparedness and response. *Public Health Rep.* 2001;116(suppl 2):9–16.

32. Yasnoff WA, Overhage JM, Humphreys BL, et al. A national agenda for public health informatics. J Public Health Manage Pract. *2001;7(6):1–21.*

33. O'Toole T. The problem of biological weapons: next steps for the nation. *Public Health Rep.* 2001;116(suppl 2):108–11.

34. Hoyo C, Hoffman I, Moser BK, Hobbs MM, Kazembe P, Krysiak RG, Cohen MS. Improving the accuracy of syndromic diagnosis of genital ulcer disease in Malawi Sexually Transmitted Disease, April 2005; 32(4):231-7.

35. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems, September 2004; 53 Suppl:159-65.

36. Dafni UG, Tsiodras S, Panagiotakos D, Gkolfinopoulou K, Kouvatseas G, Tsourti Z, Saroglou G. Algorithm for statistical detection of peaks--syndromic surveillance system for the Athens 2004 Olympic Games. September 2004; 53 Suppl:86-94.

37. Gesteland P, Gardner R, Tsui F, Espino J, Rolfs R, James B, Chapman W, Moore A, Wagner M Automated Syndromic Surveillance for the 2002 Winter Olympics, Nov/Dec 2003; Volume 10 number 6

38. Tsui F, Wagner M, Dato V, Chang C Value of ICD-9 Coded chief complaints for Detection of epidemics” American medical informatics association, volume 9, number 6, 2002.

39. Espino JU, Wagner MM, Tsui FC, Su HD, Olszewski RT, Lie Z, Chapman W, Zeng X, Ma L, Lu ZW, Dara J The RODS Open Source Project: removing a barrier to syndromic surveillance., 2004; MedInfo 11(Pt 2):1192-6.

40. https://www.rods.pitt.edu/

41. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, Olszewski RT. Classifying Free-text Triage Chief Complaints into Syndromic Categories with Natural Language Processing. Artificial Intelligence in Medicine, 2004.

42. Wagner MM, Espino J, Tsui FC, Gesteland P, Chapman W, Ivanov O, Moore A, Wong W, Dowling J, Hutman J; Syndrome and Outbreak Detection Using Chief-Complaint Data --- Experience of the Real-Time Outbreak and Disease Surveillance Project, MMWR September 2004, 53(Suppl); 28-31

43. Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Greenko J, Steiner-Sichel L, Balter S, Karpati A, Thomas P, Phillips M, Ackelsberg J, Lee E, Leng J, Hartman J, Metzger K, Rosselli R, Weiss D. New York City Syndromic Surveillance Systems, September 2004, 53(Suppl); 23-27

44. Mikosz CA, Silva J, Black S, Gibbs G, Cardenas I. Comparison of two major emergency department-based free-text chief-complaint coding systems, MMWR September 2004, 53(Suppl); 101-105

45. Wagner M., Robinson J, Tsui F, Espino J, Hogan Design of National Retail Monitor for Public Health Surveillance JAMIA 2003.

46. Labrie J. Self-care in the new millennium: American attitudes towards maintaining personal health. Consumer Healthcare Products Association, 2001, p 76

47. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. *J Fam Pract.* 1998;47:366–9

48. Tsui F-C, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc.* 2003;10:399–408

49. Zhang J, Tsui FC, Wagner MM, Hogan WR. Detection of outbreaks from time series data using wavelet transform. AMIA Annual Symposium Proc. 2003; 748-52.

50. Stoto MA, Schonlau M, Mariano LT. Syndromic Surveillance: Is it Worth the Effort? Chance, 2004

51. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making, 2003, 3:2