NHPSS – An Automated OTC Pharmaceutical Sales Surveillance System

Xiaohui Zhang, Ph.D., Reno Fiedler, and Michael Popovich

Introduction

Development of public health surveillance systems requires multiple disciplinary knowledge and advanced technologies. In this area, Halperin and Baker (1992) provided an excellent summary on public health surveillance systems, and the Centers for Disease Control and Prevention (CDC 1998 and 2001) developed new guidelines and recommendations for evaluating public health surveillance systems. It is well recognized that Over-The-Counter Pharmaceutical Sales Surveillance (OTCPSS) have a strong relationship to the public health status. However, the development and application of automated OTC pharmaceutical sales surveillance systems has received little publicity. The main reason for this is the complexity in automating this surveillance system in supporting epidemic detection, reporting and alerts. As a response to the September 11, 2001 events and the subsequent use of Anthrax by bioterrorists as well as to ensure early phase detection of possible disease outbreaks related to bioterrorism, Scientific Technologies Corp. and the State of New Hampshire Bureau of Communicable Disease Control and Surveillance have jointly developed the NH Pharmaceutical Sales Surveillance (NHPSS). This integrated electronic system automatically analyzes the OTC pharmaceutical sale data, creates reports, and initiates alerts (if necessary) for the related epidemic syndromes with their spatial and temporal characteristics. The pilot application of NHPSS was started in December of 2002 in NH Bureau of Communicable Disease Control & Surveillance to assist in epidemic detection. This paper describes the methodologies and technologies in development of OTCPSS, and summarizes NHPSS functionalities.

NHPSS was developed as a distributed information system in supporting public health surveillance. Its architecture features a database tier, an application servers tier, and a presentation/User Interface tier. The Web browser based user access and the Internet mapping functionalities provide a convenient and secure user interface.

NHPSS provides users with time trends and automated alerting in addition to ad hoc reporting. Internet mapping presents the spatial characteristics and is capable of pinpointing unusual areas of activity.

NHPSS Methodologies: A Systematic Integration of Epi-Knowledge, Statistical Methods and Information Technology

In contrast to traditional disease surveillance systems, the NHPSS was developed for pre-clinic syndromic surveillance. Traditional disease surveillance systems work in a passive mode: health care personnel review the reportable disease cases; statistical methods are applied to check if the number of cases are above the threshold conditions; then, alerts are generated if specified criteria are met. The main disadvantage of passive surveillance systems is the time delay. By the time of detection of a diagnosis-based outbreak, many people will have already been infected, and secondary spread of the contagion would often be underway. The determination of seasonally varying threshold values (control lines) is also difficult since the disease patterns vary highly in spatial and temporal dimensions. To be better prepared for the bio-terrorism events, the challenges are to improve both the timeliness and the overall decision-making process in disease surveillance systems. NHPSS is developed for automatically identifying possible pre-clinic public health problems through OTCPSS, thus early actions could be taken, rather than waiting until many people have been infected. NHPSS integrates statistical methods (for OTC data processing) and rule-based analysis (for reasoning with epidemic knowledge), with seasonally varying

base-lines and thresholds which are derived from the historical data.

Figure 1 shows the OTCPSS system structure developed by STC. The daily OTC pharmaceutical sales data are from each store, as reported to pharmacy chain headquarters. These data are replicated in data servers at the state public health department. The data warehouse organizes the data along the logical dimensions, then, the automated data processing is performed in application servers. Finally analysis, reports and alerts (if necessary) are generated to assist the decision. The user interface is Internet browser-based. With secure access, users can browse and search the reports and maps. The built-in rule-base supports the trend analysis and unusual event detections.

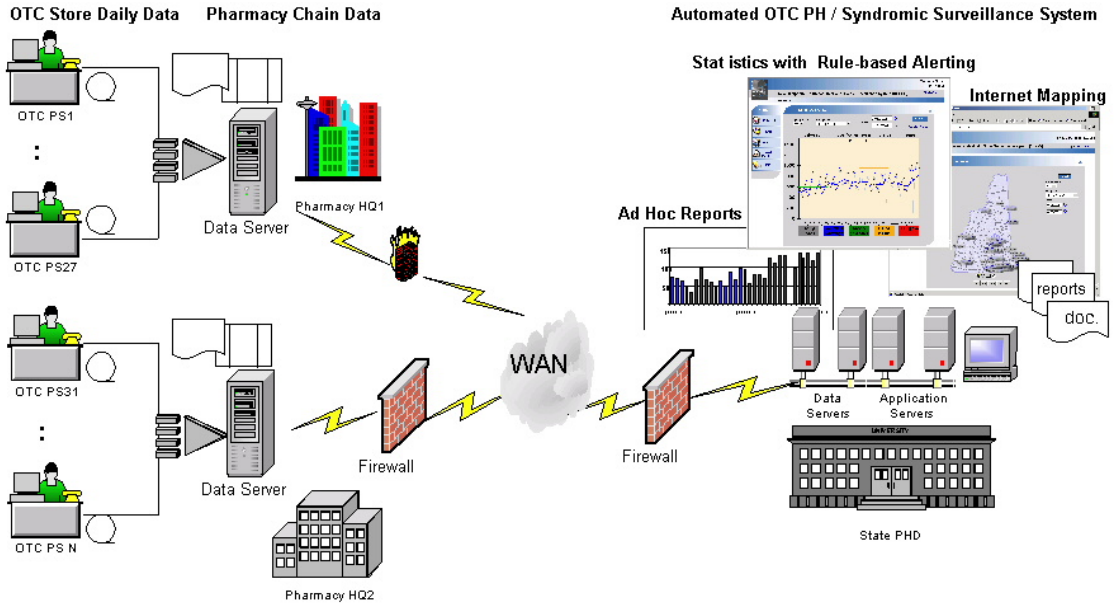

Figure 1. Diagram of the OTCPSS Structure Developed by STC

Electronic data transformation and automated data processing significantly improve the timeliness of the surveillance system. The automation of OTCPSS requires the complex integration of information technology with statistical analysis methods and abstracted epidemiological knowledge. The methodology, technology, and functionality in NHPSS development will be summarized in the following sections.

Data Warehouse: From Unprocessed Data to the Organized Information

First, the massive volumes of OTC daily sales data are categorized. Then, they are transferred into syndromic indicators, such as the Respiratory related syndrome, Gastrointestinal related syndrome, etc. The data warehouse technique is applied and the syndromic information is organized along the logical dimensions (i.e. geographical and temporal dimensions). Automated statistical processing plays an important role not only in data mining, which abstracts the spatial and temporal syndrome characteristics, but also in the generation of rule-base attributes. In NHPSS, the weekly and monthly syndromic knowledge are derived from the historic data by store, zip code area, city and state. Distributed knowledge acquisition provides the base of automated decision support in NHPSS. Figure 2 describes how the system transforms raw data to decision rule-base attributes.

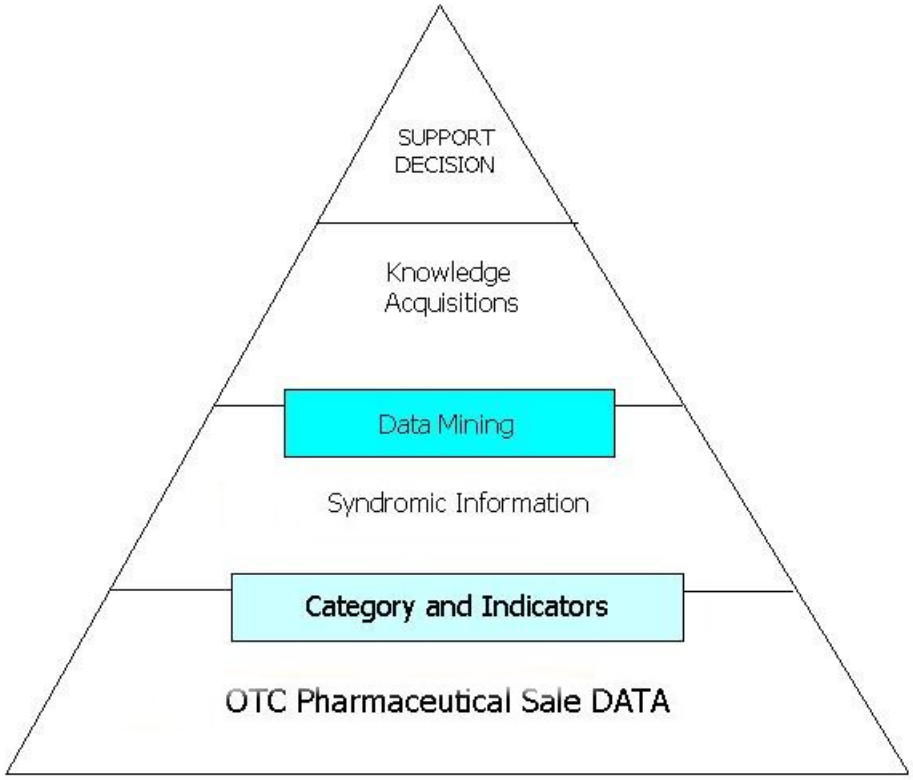

Figure 2. Diagram of the OTCPSS Knowledge Acquisition

Knowledge Acquisition and Rule-Base for Decision Making

The early detection of an epidemic event requires knowledge of both quantitative and qualitative analysis. Thus, the automation of OTCPSS needs explicit descriptions and implementations of the reasoning knowledge used in those decisions. In NHPSS, an automated statistical rule-base is developed to support event detection.

Epidemic detection requires the determination of a set of control lines or reference values. Epidemiologists then compare the data with the references to make decisions. Historically, it is a very difficult problem to set the spatial-temporal varying control lines. In NHPSS, the historical data are used to derive those distributed seasonal control lines, such as the weekly and monthly averaged daily-value at the different geographical dimensions, and different confidence intervals of the specified values. Those data-derived references are the base information for the decision-making. A rule-base is developed in NHPSS to support the pre-clinic syndrome detection. The rule-base consists of a set of logical rules attached to decision matrices. Each decision matrix has base information as the reference. It compares the decision attributes (the input data tables) with references, then interprets the data with attached rules, such as the single-point-value event, the trend identification, the mean-value shifting, clusters or cluster-drifting, which is an implementation of the epidemiologist knowledge and reasoning. Figure 3 is a diagram of this data-derived rule-base decision support.

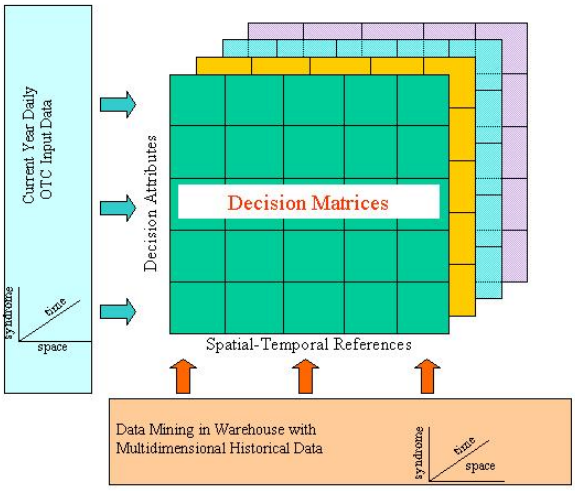

Figure 3. The Data-derived Rule-base Decision Support in NHPSS

In addition to the statistical test, a proprietary algorithm for the event detection was developed by STC. This new algorithm consists of a new filter and the rule-based detection. The application of this new algorithm to the real data has effectively detected an on-going outbreak at different geographical levels. The results showed it greatly improved the timeliness and the information yielded in assisting the assessment of the event is well beyond the capabilities of more traditional algorithms. As an example, in the studied area, the OTC daily data for an eleven (11) week period were investigated. Figure 4 is the input signals (categorized OTC daily sales in a specified place) while Figure 5 is the outputs of the filter for the above signals. The detection is then made by the rule-base. For this data set, an outbreak alert was initiated on Day 53, and the peak date was on Day 60, which are confirmed by the real case data and a non-parametric statistical test.

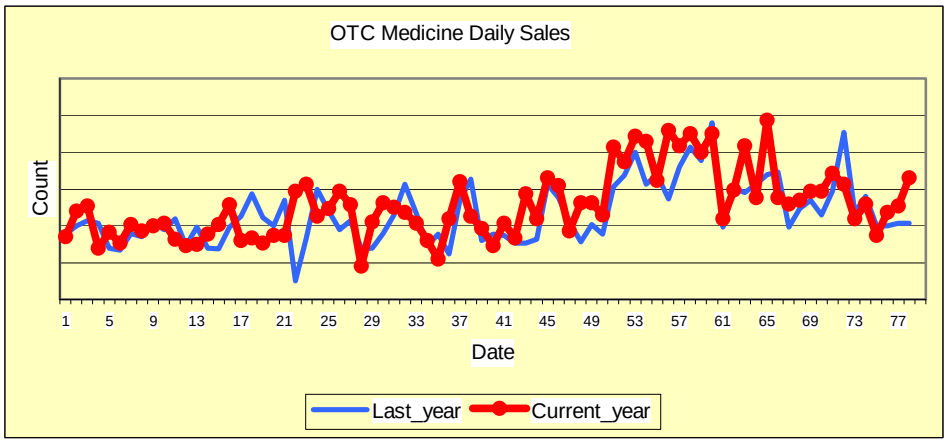

Figure 4. Current and last year's daily OTC sales

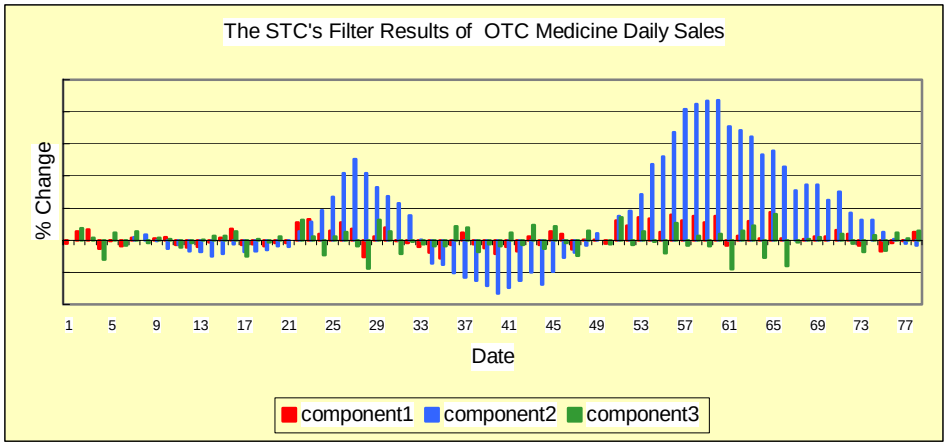

Figure 5. The filtered results for event detection

NHPSS: Application of Advanced Information Technologies

In addition to the Automated Data Processing and Rule-base integration, the application of distributed information technology is critical in NHPSS development. The implementation of NHPSS consists of the integration of Oracle Database Servers, Oracle Application Servers and ESRI ArcIMS. An important feature of NHPSS is the application of Internet GIS, which provides the spatial characteristics of a specified syndrome, with the capability to pinpoint any unusual areas.

The main functionalities in NHPSS include:

Syndromic surveillance through OTC pharmaceutical sales data:

- About 300 different Pharmaceutical items are categorized into Gastrointestinal Diseases and Respiratory Illnesses;
- The measurement of disease indicators can be the daily count of the sale packages, or the sale amount.

The automated processing for data-derived reference lines (from historical data):

- Central Line: monthly-averaged (or weekly-averaged) daily sales
- Control lines: Min., Max., N-sigma lines and Confidence Interval Upper Limits.

Analysis and Reporting in Time Dimensions:

Detailed or aggregated reporting in daily/weekly/monthly for the selected place, with capability of comparison to the historical data.

Analysis and Reporting in Geographical areas:

- Map display with alerting capability for the specified time and disease indicators.
- Pinpointing the unusual areas.

Rule-based Trend Analysis and Event Detections:

- Detection of an unusual single-point-value event by the comparison to the control lines;
- Detection of clusters and early warning of cluster-drifting;
- Detection and early warning of weekly-average shifting;
- Detection and early warning of potential trend shifting;
- Detection of starting date, peak and ending date of an event.

Figure 6 shows NHPSS hierarchical decision support from a time series alert to pinpointing the unusual areas and the local detailed reports by use of the simulation data. Note: The pilot application began in December of 2002.

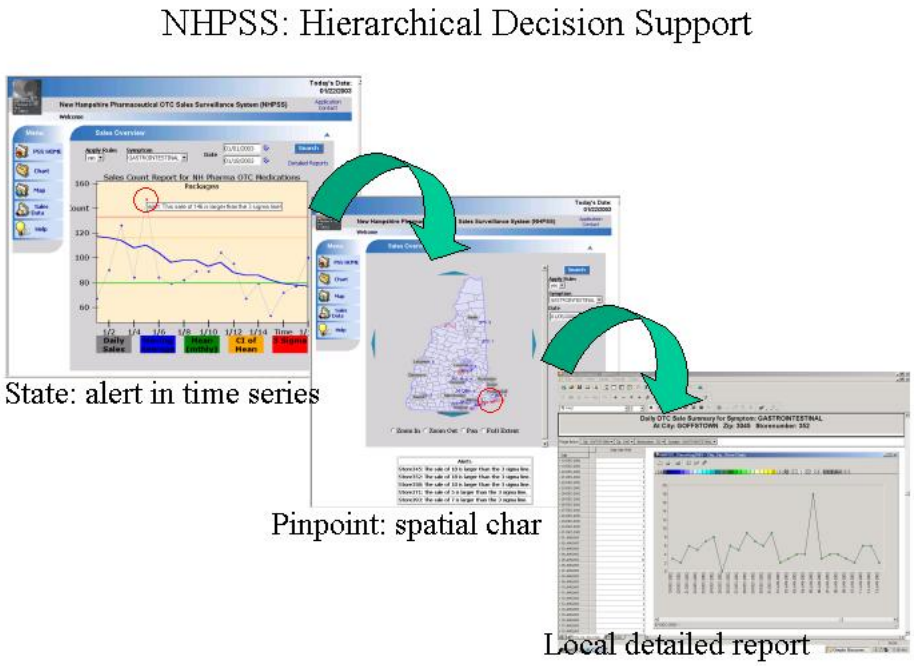

Figure 6. NHPSS for decision support at different levels

Summery and Discussions

An automated syndromic surveillance system for NH OTC pharmaceutical sales data, NHPSS, has been developed by STC for NH state government. It is a distributed information system. The pilot application has started to assist epidemic detections. The complex integration of epidemic knowledge, statistical methods and information technology has been implemented. The data-derived seasonally varying control lines from knowledge base technology make the decision base information closer to the real world. The new algorithm developed in NHPSS has greatly improved the timeliness in event detection. The pilot application of the NHPSS to the real world data yields a very promising result in pre-clinic syndromic surveillance. Finally, the automated data processing with rule-based analysis and event detection (with alerting) distinguishes NHPSS over other surveillance systems.

References

Halperin, W. and Baker, E. L. 1992, Public Health Surveillance, New York: Van Nostrand Reinhold.

CDC 1998, Guidelines for evaluating surveillance systems, MMWR May 6, 1998; 37(S-5).

CDC 2001, Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group, MMWR July 27, 2001; 50(RR-13).

Scientific Technologies Corporation
4400 E. Broadway Blvd.,
Tucson, AZ 85711