**RC23368 (W0410-060) October 8, 2004**
**Computer Science**

# IBM Research Report

## An Evaluation of Over-The-Counter Medication Sales for Syndromic Surveillance

**Murray Campbell, Chung-Sheng Li, Charu Aggarwal, Milind Naphade, Kun-Lung Wu, Tong Zhang**
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

**Research Division**
**Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich**

# An Evaluation of Over-The-Counter Medication Sales for Syndromic Surveillance

Murray Campbell Chung-Sheng Li Charu Aggarwal Milind Naphade
Kun-Lung Wu Tong Zhang

IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
USA

## Abstract

*Early and reliable detection of disease outbreaks is an important problem for public health. Syndromic surveillance systems use pre-diagnostic data sources to attempt to improve the timeliness of outbreak detection. This paper describes a number of approaches to evaluating the utility of data sources in a syndromic surveillance context. We show that there is some evidence that sales of over-the-counter medications have value for syndromic surveillance.*

## 1 Introduction

Syndromic surveillance refers to the use of pre-diagnostic health-related data for early detection of disease outbreaks. With recent concern over the threat of bioterrorism, as well as the appearance of new disease threats (e.g., SARS), syndromic surveillance is being looked to as a means to improve the timeliness of public health surveillance.

The development of a useful syndromic surveillance system depends in part on the identification of data sources that have value in predicting disease outbreaks. This paper will focus on methods for assessing the value of data sources for predicting disease outbreaks. We will examine a number of different approaches that use retrospective analysis to evaluate data sources.

A frequently cited example of a data source that is presumed to be useful for syndromic surveillance is the sale of over-the-counter (OTC) medications. We will apply our evaluation approaches to a large, multi-year, multi-city data set and show that there is some evidence that OTC medication sales may be useful for syndromic surveillance.

## 2 Background and Related Work

Syndromic surveillance (also referred to in the literature as early detection of disease outbreaks, pre-diagnosis surveillance, non-traditional surveillance, enhanced surveillance, non-traditional surveillance, and disease early warning systems) has received substantial interest recently, especially after Sept. 11, 2001 [3, 5, 9, 12, 13, 14, 15].

A number of studies have been devoted to investigating various data sources, such as the text and the ICD-9 diagnosis code of the chief complaints from emergency department [1, 2, 6, 11], 911 calls [4], and over-the-counter (OTC) drug sales [8].

There are at least three different classes of approaches to evaluating the utility of a data sources for syndromic surveillance. The first approach is based on the measuring the correlation between a target data source and a gold standard (diagnostic) data source [16]. A second approach is to use the target data source to better predict values in the gold standard data source. A third option is to identify "events" (i.e., disease outbreaks) in a gold standard data source, and assess the timeliness of alarms produced by a detection algorithm operating on the target data source. The tradeoff between timeliness and false alarms can be assessed using the AMOC approach [7].

## 3 Data

There are two data sets that will be used in this study. The first, which we will call OTC, is a weekly summary of unit sales of upper respiratory over-the-counter medication sales for ten cities (Baltimore/Washington, Charlotte, Chicago, Dallas, Milwaukee, New York, Norfolk, Orlando, Pittsburgh, and Seattle) for a three-year period (2000-2002). The first data point is for the week ending on 1/9/2000, and the last data point is for the week ending 12/29/2002. For each city, sales are reported in eight categories: four types

(Cold, Allergy, Cough, and Sinus), and two target groups for each type (Adult and Pediatric).

The second data set, which we will call CL, consists of anonymized medical insurance claims records. The records are from the same ten cities as for OTC, and cover the same three-year period. Each record consists of a unique (anonymized) patient identifier, a date of service, up to four ICD-9 (diagnosis) codes, and a city name. There are a total of about 22.5 million records. The ICD-9 codes were chosen by the data provider, Surveillance Data, Inc., to be relevant to upper respiratory infections. The number of insurance claims were aggregated by city to weekly totals aligned with the OTC data.

For the purposes of this study, the OTC data set is the target data source, i.e., OTC will be assessed for value in syndromic surveillance. CL is the gold standard data source, as it contains diagnostic information about actual disease.

# 4 Approaches

## 4.1 Lead-Lag Correlation Analysis

One approach to evaluating a data source for syndromic surveillance is to conduct a lead-lag correlation analysis on the data source with respect to a gold standard data source. This consists of computing the correlation between the two time series for a range of lead-lag times, and identifying the lead-lag time at which the correlation is maximized. It can be useful to remove trends before analyzing.

Although a correlation analysis can give a global view of the lead time of a target data source, syndromic surveillance is typically more interested in the lead time prior to increasing levels of disease. This suggests an alternative approach where a correlation analysis is performed on a number of shorter time segments that contain the initial stages of disease outbreaks.

In Section 5.1 we will apply this method to the data sets described in Section 3, and assess the value of OTC data for syndromic surveillance.

## 4.2 Regression Test of Predictive Ability

This section describes another approach to evaluating the usefulness of a target data source by posing it as a prediction problem. More specifically, we are interested in predicting certain quantities associated with the gold standard data source, and want to see whether by including the target data, we are able to make better predictions.

This approach can be generally regarded as time-series forecasting. If we can forecast a quantity $A$ more accurately using a quantity $B$ under a certain metric, then we say that $B$ contains useful information for predicting $A$.

We now give a general description of this approach. Assume that the quantity of interests is presented sequentially as a time-series

$$\{Y\} = \{\cdots, Y_0, Y_1, \cdots, Y_t, \cdots\}.$$

We want to predict the future values of this time-series based on some side-information (which may includes the historical values of $Y$ we observed so-far), represented as another time-series of vectors:

$$\{\mathbf{X}\} = \{\cdots, \mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots\}.$$

Each $Y_t$ is a real-valued number, observed at time $t$, which we are interested in. Each $\mathbf{X}_t$ is a real-vector, which encodes all of the side information that we hope are useful for predicting the $\{Y\}$ series.

To this end, we assume that at each time $t$, based on the current side-information $X_t$, we would like to predict $Y_{t+f}$, which is the value of the $Y$ series $f$-steps in the future (where $f > 0$ is an integer). We assume that the predictor $p_f(\mathbf{X}_t)$ has a linear form as

$$Y_{t+f} \approx p_f(\mathbf{X}_t) = \mathbf{w}_f^T \mathbf{X}_t,$$

where $\mathbf{w}_f$ is a weight vector (parameter of our model) that characterizes the predictor $p_f$. The parameter $\mathbf{w}_f$ can be estimated from the data (as we will describe later).

Given a predictor, represented as a weight vector $\mathbf{w}$, we can measure its quality using a certain figure of merit. In this study, we employ the commonly used least-squares error criterion, defined as

$$R_f(\mathbf{w}, [T_1, T_2]) = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} (\mathbf{w}^T \mathbf{X}_t - Y_{t+f})^2.$$

The number $R_f(\mathbf{w}, [T_1, T_2])$ measures in the interval $[T_1, T_2]$, how well we can predict from $X$ the sequence $Y$ $f$-steps in advance with the weight vector $\mathbf{w}$.

The weight vector can be estimated from the historical data using least-squares regression:

$$\hat{\mathbf{w}}_{f,T} = \arg\min_{\mathbf{w}} \sum_{t=1}^{T-f} (\mathbf{w}^T \mathbf{X}_t - Y_{t+f})^2. \qquad (1)$$

Now assume that we observe the sequences $\mathbf{X}$ and $Y$, up to some point $T$. To check how useful is $\mathbf{X}$ for predicting $Y$, we divide the time period into $K$ consecutive blocks (for simplicity, assume that $T$ is divisible by $K$): $I_j = [T_j, T_{j+1}]$ for $j = 0, \ldots, K-1$, where $T_j = jT/K$. Now we can use a single number

$$r_f(\mathbf{X}, Y) = \frac{1}{K} \sum_{j=1}^{K-1} R_f(\hat{\mathbf{w}}_{f,T_j}, [T_j, T_{j+1}]) \qquad (2)$$

to measure the usefulness of $\mathbf{X}$ for predicting $Y$ ($f$-steps in the future). That is, we train a predictor $\hat{\mathbf{w}}_{f,T_j}$ using least squares regression (1) with data observed up to $jT_0$, and then test on data from $jT_0$ to $T_{j+1}$, for $j = 0, \ldots, K-1$, and then average the results. The smaller $r_f(\mathbf{X}, Y)$ is, the more useful $\mathbf{X}$ is for predicting $Y$. Therefore using (2), we can compare the usefulness of different side informations $\mathbf{X}$ and $\mathbf{X}'$.

In Section 5.2, we compute the corresponding $r_f(\mathbf{X}, Y)$ numbers with and without including the OTC data in the side information $\mathbf{X}$. Our results suggest the usefulness of the OTC data in public health surveillance.

### 4.3 Detection-Based Approaches

For the detection-based approaches we assume that disease outbreak events are labeled in the gold standard data set, and an outbreak detection algorithm operates on either the the target data set or the gold standard data set. Using the AMOC approach, we are able to assess the lead time provided by the target data source over a range of practical false alarm rates.

#### 4.3.1 Supervised Algorithm for Outbreak detection in OTC data

The supervised outbreak detection algorithm utilized the previously supplied data in order to determine various aspects of the algorithm. The supervised algorithm required a number of components in order to perform the detection:

(1) Determination of features to be used, and the proper way to combine channels.

(2) Creation of streams of anomalies.

(3) Conversion of the anomaly streams into the alarm level using the information from (1).

This supervision was done in two forms:

(1) Feature Selection: Since multiple channels of information were available, which channels provided the greatest level of connection between the channels and actual outbreaks?

(2) Combination of Multiple Channels: How do we combine the signals from multiple channels in order to create one integrated alarm level which was most effective for detecting the outbreak?

In order to perform feature selection, we used the same OTC data set (provided by SDI) as described in the other sections. The first step was to determine which of the channels were most discriminatory for the purpose of distinguishing the biological outbreak from the background noise.

Let us assume that for each site $i$, the value indicating the channel specific information (absentee behavior, phone calls, pharmacy buying behavior) at time $t$ is denoted by $y(i,t)$. The first step was to convert the data into statistical deviation levels which could be compared across different features. Thus, each stream of data was converted into a statistical stream of numbers indicating the deviation level with respect to the prior window of behavior of width $W$. The statistical deviation value for a given stream $i$ at time $t$ was denoted by $z(i,t)$. The value of $z(i,t)$ was found by first fitting the prior window of with $W$ with the polynomial function $f(t)$. The deviation value at time $t0$ was then defined as follows:

$$s(i) = \sqrt{\sum_{t=t0-W}^{t0} (f(t) - y(i,t))^2/(W-1)} \qquad (3)$$

The value of W used was based on the last 16 reports. This statistical deviation is also referred to as the $z$-number. This value provides an idea of how far the stream of data deviates from the normal behavior and gives an intuitive understanding of the level of anomaly at a given tick. Then, the statistical deviation $z(i,t0)$ at time $t0$ is denoted by:

$$z(i,t0) = (f(t0) - y(i,t0))/s(i) \qquad (4)$$

These alarm values could be used in order to determine the value of each channel in the training data. A particular channel was found to be useful when this value was found to be larger than a pre-defined threshold of 1.5. For example, by using this technique we were able to eliminate the allergy channel for the purpose of detection of the flu infections. For example, this behavior was illustrated by the allergy channel in the OTC training data. We have also illustrated the AMOC curve for the allergy channel in the same figure. We note that the AMOC curve for the allergy channel was particularly poor, because it seemed to be uncorrelated to the seasonal outbreaks in the data.

Once these features were selected, they could be used on the test data for computing the statistical deviation values using the same methodology as discussed above. Thus, a separate signal was obtained from stream. The next step was to combine the deviation values from the different sites and channels to create one composite signal. A supervised training process was utilized to determine the optimal functional form for the test data. This was achieved my finding the composition which maximized the area under the AMOC curve.

Once each channel had been converted into a single composite signal, they need to be combined to create a combination signature. For example, let $q1(t)$, $q2(t)$ and $q3(t)$ be the signatures obtained from three different channels. The combination signature was defined as the expression:

$$C(t) = c1 \cdot q1(t) + c2 \cdot q2(t) + c3 \cdot q3(t) \qquad (5)$$

Here $c1$, $c2$ and $c3$ were coefficients which were also determined by minimizing the latency of detection on the training data. As a normalization condition, it is assumed that

the coefficients satisfy the following condition for the constant $C'$:

$$c1^2 + c2^2 + c3^2 = C' \quad (6)$$

It is necessary to use the above condition for scaling purposes. In order to determine the optimal alarm we found values of $c_1$, $c_2$, and $c_3$, which optimized the area under the AMOC curve. This provides the combination signature.

#### 4.3.2 Modified Holt-Winters forecaster

One of the unsupervised detectors used was a modified Holt-Winter forecaster [10]. The forecaster generate a z-value for each tick of a data channel, representing the deviation of observed data from the predicted one. A z-value is computed as follows:

$$z = (\Delta - \mu)/\sigma,$$

where $\Delta$ is the difference between observed and predicted data, and $\mu$ and $\sigma$ are the mean and standard deviation, respectively, of these $\Delta$ differences in the past.

A Holt-Winters forecaster assumes that a time series, $X_1, \cdots, X_N$, can be modeled in terms of three key components: the average $\overline{X}_N$, the trend $T_N$ and the daily seasonality factors $F_{N-D+1}, \cdots, F_N$, where $D$ is the number of days in the week for which there are observed data. The average is the exponentially smoothed level value of all the time series values. The trend is the exponentially smoothed slope of all the $N$ time series values. The daily seasonality factors are exponentially smoothed values reflecting the deviation from linearity attributable to the different days of the week. The seasonality factors can have either a multiplicative or additive effect. In our implementation, we chose the additive variant. A Holt-Winters forecaster attempts to accurately capture these three key components of a time series. It can deal with special events, such as holidays or special days where data are missing.

#### 4.3.3 Forecasting based on Multi-channel Regression

A simple prediction strategy that can combine single and multi-channel prediction is to set up the problem as a linear regression. As usual, the deviation of the actual value from the predicted value as a measure of abnormality. We set up a system of linear equations as shown below.

Let the observation stream of a single channel from among the multiple OTC sales channels be $[y_1, \ldots, y_M]$. Consider using the past $J$ observations to derive the regression parameters while using the past $K$ samples for actually predicting the $K + 1^{th}$ observation. The number of variables to be estimated from the past $J$ samples is $K$.

$$\begin{bmatrix} y_{M-1} \\ y_{M-2-1} \\ \cdots \\ y_{M-J-1} \end{bmatrix} = \begin{bmatrix} y_{M-2} & \cdots & y_{M-K-1} \\ y_{M-3} & \cdots & y_{M-K-2} \\ \cdots & \cdots & \cdots \\ y_{M-J-2} & \cdots & y_{M-K-J-1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_K \end{bmatrix} \quad (7)$$

or using matrix notation:

$$Y = A_y W, \quad (8)$$

With this overdetermined system of equations ($J >$K) we then calculate the least squares fit to this as shown in Eq 9:

$$W = (A_y^t A_y)^{-1} A_y^t Y \quad (9)$$

Assuming linear independence among columns of matrix $A$, $A^t A$ is non singular and the generalized inverse $(A^t A)^{-1}$ exists. We calculate the weight vector $W$ after every update. Thus for each observation $y_M$ we calculate the prediction $aW$, $a$ being a row vector $[y_{M-1} y_{M-2-1} \ldots y_{M-J-1}]$. If the residual between the actual value and the predicted value is positive we use this difference as a measure of abnormality and probability of an outbreak. Equation 7 can be extended to make the prediction based on multiple channels. For example the matrix $A$ can be created by combining multiple channels. Equation 10 shows past samples from two channels $[y_1, \ldots, y_M]$ and $[x_1, \ldots, x_M]$ being used to predict the current observation of channel Y.

$$Y = [A_y A_x] \begin{bmatrix} W_y \\ W_x \end{bmatrix} \quad (10)$$

Using the above formulation we can predict the current value of sales of any of the OTC channel based on values of sales in the same channel as well as based on values of sales in additional channels.

## 5 Experiments

### 5.1 Lead-Lag Correlation Analysis of OTC Data

The lead-lag correlation analysis approach requires us, for each city, to compute the correlations corresponding to various possible lead-lag times. In Figure 5.1, we examine offsets ranging from five weeks prior to five weeks after. The ten solid lines are the correlation values for each of the ten cities. The dashed line is the mean of those values. The peak correlation is between one and two weeks leading, i.e., OTC leading CL by one to two weeks. If a quadratic is fitted to the dashed line, the maximum is at 1.7 weeks.

The provides evidence, albeit somewhat weak, that OTC leads CL and may have value for syndromic surveillance. Clearly there is a wide discrepancy on the correlation between OTC and CL across the different cities, and this needs further investigation.

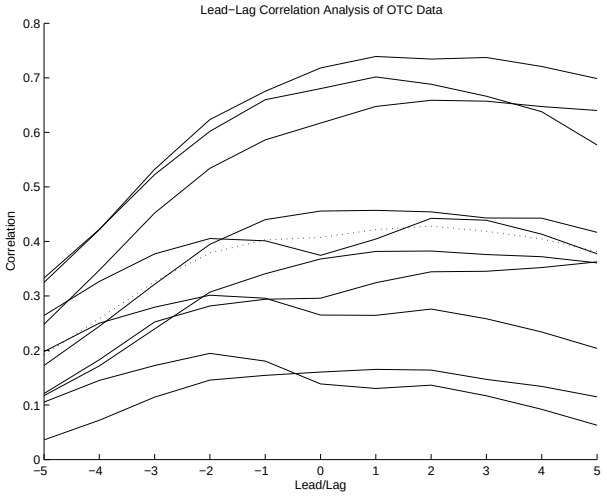


**Figure 1. Lead-Lag correlation analysis experiment**

## 5.2 Regression Test of the Predicative Value of OTC

We study the usefulness of OTC for predicting insurance claims using the approach described in Section 4.2. Since the OTC data are weekly based, we shall form the time series on a weekly basis. In particular, we convert the insurance data into weekly data aligned with the OTC data.

In this experiment, we consider different cities separately. That is, we do not consider possible inter-city correlations. For each city, we let $OTC_t$ be the total number of OTC sales in week $t$, and $CL_t$ be the number of insurance claims in week $t$. Since in public health surveillance, we are mostly interested in sudden outbreaks of diseases, we are interested in the log-ratio of the number of insurance claims in consecutive weeks. That is, at week $t$, the $Y$ variable is given by

$$Y_t = \log_2(CL_t/CL_{t-1}).$$

One may also use other quantities, such as whether the insurance claims next week is higher than this week by a certain amount (or whether $Y_t$ is larger than some threshold).

We consider a few possible side information $\mathbf{X}$, which we list below.

- $\mathbf{X}^1$: Using constant side information: $\mathbf{X}^1_t \equiv [1]$. This leads to a predicator that predict $Y_t$ using its historic mean.
- $\mathbf{X}^{CL}$: In addition to the above, we also include historical observations of the insurance claim data itself (the log ratio of the current number of claims over the claims of the previous week) as side-information: $\mathbf{X}^{CL}_t = [Y_t, 1]$.
- $\mathbf{X}^{OTC}_t$: We include the constant one and the OTC data into the side-information:

$$\mathbf{X}^{OTC}_t = [\log_2(OTC_t/OTC_{t-1}), 1].$$

- $\mathbf{X}^{CL-OTC}_t$: We include all of the above quantities into the side-information:

$$\mathbf{X}^{CL-OTC}_t = [\log_2(OTC_t/OTC_{t-1}), Y_t, 1].$$

Since this framework is quite flexible, various other configurations can also be studied. For our purpose, we are able to make interesting observations from this particular configuration. Variations will lead to similar results.

Applying the notation in Section 4.2, for each city, we divide the time series into $K = 20$ blocks, and compute the $r_f(\mathbf{X}, Y)$ number in (2) for $f = 1, 2$ and each side information listed above. We then average the results over the ten cities, and report the averaged numbers in Table 1. From the table, we can see that the OTC data has a small predicative power for the insurance claims data CL. One may also do an experiment in the reverse order (that is, use historical CL data to predict the future OTC sales). In this case, for $f = 1$, the predictive performance for OTC sales, measured by the $r_f$ value, degrades from 0.0217 (without CL in the side-information) to 0.0221 (with historical CL data in the side-information). Therefore these experiments provide some evidence suggesting that OTC changes precede CL changes.

| | $\mathbf{X}^1$ | $\mathbf{X}^{CL}$ | $\mathbf{X}^{OTC}$ | $\mathbf{X}^{CL-OTC}$ |
|---|---|---|---|---|
| $f = 1$ | 0.0287 | 0.0265 | 0.0285 | 0.0261 |
| $f = 2$ | 0.0287 | 0.0291 | 0.0280 | 0.0287 |

**Table 1. Averaged $r_f(\mathbf{X}, Y)$ numbers over ten cities**

Although effects shown in Table 1 are relatively small, we believe they are still indicative statistically. Since we average our results over ten cities, we may also check the variation over different cities. In particular, in seven out of ten cities, $r_1(\mathbf{X}^{OTC}, Y)$ is smaller than $r_1(\mathbf{X}^1, Y)$; also in seven out of ten cities, $r_2(\mathbf{X}^{OTC}, Y)$ is smaller than $r_2(\mathbf{X}^1, Y)$. This comparison is consistent with results in Table 1, and justifies from a slightly different point of view that statistically, the OTC data is (weakly) useful for predicting future insurance claims.

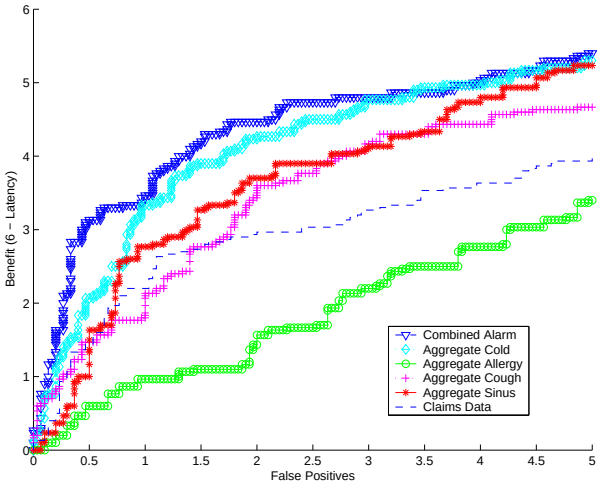


**Figure 2.** The AMOC curves generated by the Supervised method illustrate that various OTC categories are more timely than claims.

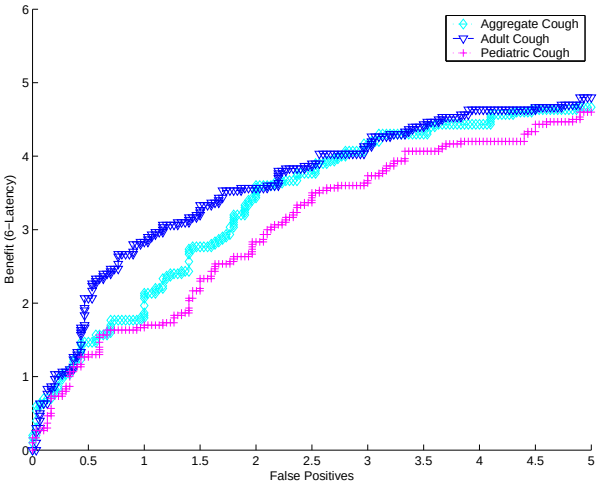


**Figure 3.** The AMOC curves generated by the Supervised Method illustrate that there is differentiation between Adult and Pediatric cough medication sales.

## 5.3 Results From Detection-Based Approaches

### 5.3.1 Supervised Method

Once the features have been selected, and the proper way for construction of the combination signature was determined, the actual alarm level construction on the data was straightforward. The deviation values for the data were computed in an exactly identical way to the training data, and the combination was created to output the corresponding alarm levels at each tick. In Figure 5.3.1, we have illustrated the behavior of the detection algorithms. Once interesting observation was that the OTC data was always more effective than the claims data. In fact, in most cases, the OTC data acted as a "leading indicator" over the claims data. It is also interesting to note that the adult and pediatric data illustrated differential behavior in terms of the speed and quality of the detection. An example of this is illustrated in Figure 5.3.1.

### 5.3.2 Modified Holt-Winters forecaster

Even though the OTC data were weekly data, the detector treated them as daily data and assumed that there were 3 days in a week. It used the past 6 OTC data points to predict the next OTC sale.

While there was some variability across different categories of OTC medication sales, over a wide range of false alarm rates the Holt-Winters forecaster showed a two week lead time for OTC over Claims. Sinus medication sales were observed to have the best lead times overall.

### 5.3.3 Forecasting based on Multi-channel Regression

Using the OTC data we experimented with different values of $J$ and $K$ (see Section 4.3.3 for single and multichannel prediction based outbreak detection. Based on our experiments we found that sales of adult drugs were more informative about the outbreakss and had a lead time of between 2 and 3 weeks over claims. We also found encouraging empirical evidence that the use of multiple channels resulted in a better lead time for predicting outbreaks over single channel prediction. Figure 4 shows the AMOC curve using the adult cold channel for predicting outbreaks. It also shows the benefit of using adult cold and adult cough to predict adult cold sales and use the deviation to detect outbreaks although this benefit is evident only for small values of false alarms as seen in the AMOC curve

## 6 Conclusions and Future Work

We have shown a number of different approaches to assessing the value of a data source for syndromic surveillance, and evaluated over-the-counter medication sales using these approaches. The appears to be evidence from each of these approaches that OTC medication sales are a leading indicator for disease outbreaks.

There are a number of limitations in this study. The data sets were aggregated weekly, which reduces the precision regarding estimates of the timeliness of OTC. This type of study should be repeated with daily data. The detection-based experiments identified only those outbreaks that occurred at the beginning of the seasonal rise in respiratory disease. A more careful study could examine finer grain

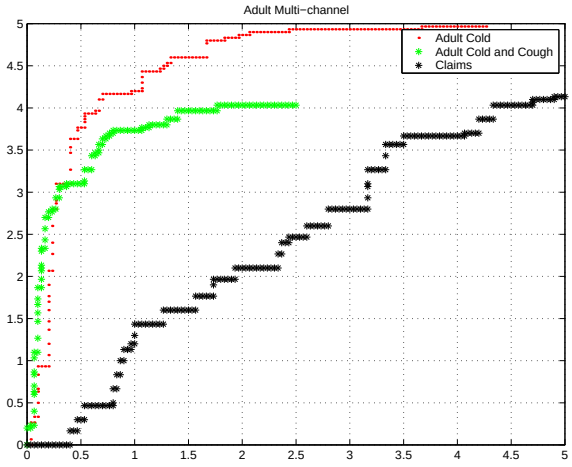


**Figure 4.** The Adult Cold sales were found to be the best indicator for the outbreaks with $J = 15, K = 2$ and $J = 20, K = 1$ respectively for single channel and multi-channel prediction. The Adult Cold and Cough sales were used in the two channel prediction.

disease outbreaks, preferably those that have been studied and verified by public health. This study was retrospective, looking only at historical data. A prospective study, using the target data source to predict disease outbreak in real time, would provide greater confidence in the conclusions in this paper.

## 7 Acknowledgments

We would like to thank Andrew Kress of Surveillance Data, Inc. for supplying the data used in this study. This work is supported by the Air Force Research Laboratory (AFRL)/Defense Advanced Research Projects Agency (DARPA) under AFRL Contract No. F30602-01-C-0184. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the AFRL or DARPA.

## References

[1] E. M. Begier, D. Sockwell, L. M. Branch, J. O. Davies-Cole, L. H. Jones, L. Edwards, J. A. Casani, and D. Blythe. The national capitol region's emergency department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? *Emerging Infectious Disease*, 9(3):393–396, Mar. 2003.

[2] A. J. Beitel, K. L. Olson, B. Y. Reis, and K. D. Mandl. Use of emergency department chief complaint and diagnostic codes for identifying repiratory illness in a pediatric population. *Pediatric Emergency Care*, 20(6):355–360, Jun. 2004.

[3] J. W. Buehler, R. L. Berkelman, D. M. Hartley, and C. J. Peters. Syndromic surveillance and bioterrorism-related epidemics. *Emerging Infectious Diseases*, 9(10):1197–1204, Oct. 2003.

[4] M. R. Dockrey, L. J. Trigg, and W. B. Lober. An information systems for 911 dispatch monitoring system and analysis. In *Proceeding of the AMIA Sympoisum*, page 1008, 2002.

[5] J. S. Duchin. Epidemiological response to syndromic surveillance signals. *Journal of Urban Health*, 80(2):i115–i116, 2003.

[6] J. U. Espino and M. M. Wagner. Accuracy of icd-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. In *Proceedings AMIA Symposium*, pages 164–168, 2001.

[7] T. Fawcett and F. Provost. Activity monitoring: Noticing interesting changes in behavior. In Chaudhuri and Madigan, editors, *Proceedings on the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 53–62, San Diego, CA, 1999.

[8] A. Goldenberg, G. Shmueli, R. A. Caruana, and S. E. Fienberg. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):5237–5240, Apr. 2002.

[9] T. Goodwin and E. Noji. Syndromic surveillance. *European Journal of Emergency Medicine*, 11(1):1–2, Feb. 2004.

[10] C. Granger and P. Newbold. *Forecasting Economic Time Series*. Academic Press, 1977.

[11] J. Greenko, F. Mostashari, A. Fine, and M. Layton. Clinical evaluation of the emergency medical services (ems) ambulance dispatch-based syndromic surveillance system, new york city. *Journal of Urban Health*, 80(2):i50–i56, 2003.

[12] F. Mostashari and J. Hartman. Syndromic surveillance: A local perspective. *Journal of Urban Health*, 80(2):i1–i7, 2003.

[13] J. A. Pavlin. Investigation of disease outbreaks detected by syndromic surveillance systems. *Journal of Urban Health*, 80(2):i107–i114, 2003.

[14] D. M. Sosin. Syndromic surveillance: The case for skillful investment. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 1(4):247–253, 2003.

[15] M. M. Wagner, F. C. Tsui, J. U. Espino, V. M. Dato, D. F. Sittig, R. A. Caruana, L. F. McGinnis, D. W. Deerfi eld, M. J. Druzdzel, and D. B. Fridsma. The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management Practice*, 7(6):51–59, Nov. 2001.

[16] R. Welliver, J. Cherry, K. Boyer, J. Deseda-Tous, P. Krause, J. Dudley, R. Murray, W. Wingert, J. Champion, and G. Freeman. Sales of nonprescription cold remedies: a unique method of influenza surveillance. *Pediatric Research*, 13:1015–1017, 1979.