Basic Models for Mapping Prescription Drug Data

Kennon R. Copeland and A. Elizabeth Allen
IMS Health, 660 W. Germantown Pike, Plymouth Meeting, PA 19462

Abstract

Geographic patterns of diseases across time are of critical interest for public health monitoring and policy. Availability of information from which to derive estimates of disease patterns, however, is generally very limited, either in terms of geographic or population coverage, or both. By contrast, availability of information on pharmaceutical drug dispensing is very robust. Pharmaceutical drug information, although not directly indicative of disease, may provide useful surrogate information about disease patterns. The nature of pharmaceutical drug information presents new challenges in terms of developing appropriate estimates of geographic distributions. This paper discusses pharmaceutical drug information obtained from retail pharmacies, presents a set of basic models for consideration, examines the application of basic models for one pharmaceutical product, and outlines future research.

Keywords: Syndromic Surveillance, Disease Outbreak Detection, CuSUM, SaTScan

1. Overview

Early accurate detection of disease outbreaks is a critical problem for public health monitoring and, in the post 9/11 environment, homeland security. In addition to disease outbreak detection, there are public health concerns associated with early detection of changes in disease prevalence for chronic diseases.

Rolka (2004) identified eight factors upon which successful detection relies, among them:

*Initial detection – find an event as early as possible
*Quantification – how many people are ill?
*Localization – where is it taking place?

The current environment of data collection and availability of diseases information has a variety of issues, including:

*Lack of comprehensive disease and/or diagnosis information
*Low or incomplete coverage of incidents
*Reporting lags between incidence and availability for detection

These issues are contrary to the factors associated with successful disease outbreak and increased disease prevalence detection.

Syndromic surveillance (see, e.g., Stoto, et al., 2004) offers the opportunity to address the lack of disease specific information by identifying correlates of a disease and monitoring these correlates for detection purposes.

One such correlate is pharmaceutical drug usage, for which there exists timely data with high coverage rates. Prescription dispensing of pharmaceutical drugs provide an indirect indicator of disease incidence. An underlying question associated with the use of prescription drug information is the correlation between a pharmaceutical drug and a particular disease, as drugs may be used for multiple diseases, or may not be used by all persons who have the disease.

2. Description of Data Source

IMS collects prescription information from roughly 34,000 retail pharmacies nationwide on a weekly basis. This sample represents approximately 67% of retail pharmacies and 73% of retail prescription volume, and is geographically spread throughout the U.S. The reporting week is Saturday through Friday. Prescription information provided to IMS is that recorded within pharmacy software systems as part of regular prescription management conducted by pharmacies. Thus, there is an incentive for complete and accurate reporting by pharmacies.

IMS utilizes a geo-spatial ratio estimation approach to estimating Rx activity within individual nonsample pharmacies, with weights applied to sample pharmacies based upon the relative pharmaceutical purchase volume of the nonsample to sample pharmacies and inversely proportional to the distance between sample pharmacies and the nonsample pharmacy. The methodology yields prescriber level estimated prescription volume at the product/form/strength level, which can be summed to any geographic level from zip code to national level. Estimates from the sample are reported on a weekly basis, 10 days following the week of interest.

In addition, IMS collects prescription information on a daily basis from a subsample of the weekly pharmacies, constituting roughly 15% of retail prescription volume. Data from these pharmacies are currently used to generate national level estimates of prescription volume at the product/form/strength level at noon the second day following the day of interest.

3. Basic Models

Two methods commonly used to analyze disease measures and detect outbreaks are CuSUM and SaTScan.

CuSUM (see, e.g., Raubertas, 1989) is a sequential data analysis approach that uses the sum of deviations of observed from expected to detect changes in distribution across time. CUSUM methods are designed to detect sudden changes in the mean value of a quantity of interest methods, and are used to monitor disease counts in geographic areas of interest.

SaTScan (Kulldorff, 2005) is a cluster detection software that uses Poisson, Bernoulli, and space-time permutation models to group observations into homogeneous clusters. The software detects changes in distributions and identifies clusters across space, time, or space-time.

4. Case Study

The disease commonly used to evaluate the performance of data sources and models is influenza. CDC provides weekly reports on incidence based upon approximately 1,000 Sentinel health care providers surveyed each week from a variety of healthcare settings (CDC, 2005a). The Sentinel providers report actual number of Influenza Like Illness (ILI) visits. ILI is defined as a fever of >100° with either cough or sore throat. The reporting week is Sunday through Saturday.

Sentinel provider data are weighted to account for differences in demographic distribution between the sample and full populations, to yield national level incidence rate estimates. The estimates and analysis are made available a week following the reporting week.

There are time series data on ILI estimates at both the national and regional level for a number of years upon which historical comparisons can be made.

The prescription drug Tamiflu is approved for use in early treatment of influenza and, therefore, prescription volumes for Tamiflu may serve as a usable proxy for influenza incidence.

We examined Tamiflu Rx patterns for the 2003-2004 and 2004-2005 flu seasons. Tamiflu weekly Rx estimates at the national, regional, and state level were obtained using data from the IMS Xponent™ prescription database. These estimates were compared to reports of influenza from CDC (CDC, 2005b; CDC, 2005c), to examine the outbreak detection features of the prescription data.

For the Rx data, estimated prescription volumes were divided by 2004 Census Bureau population estimates (U.S. Census Bureau, 2005) to obtain usage rates per 100,000 population. The CDC ILI data are reported as percent of patient visits for ILI. While these two measures are not directly comparable, they do allow for comparisons of timing, trends, and magnitude.

5. Results

5.1 National Level

National level estimates of Tamiflu usage appear comparable to national level estimates of ILI visits in terms of timing, trends, and magnitude for both the 2003-2004 and 2004-2005 influenza seasons (Figures 1 and 2).

As seen in the Figures, both series show the influenza outbreak in 2003-2004 occurred earlier (Nov-Dec vs. Jan-Feb), and was more severe (40%-50% greater measures) than in 2004-2005. Although the trends and magnitudes for the two series are consistent, differences in the Rx and ILI measures keep the series from completely overlaying. In addition, there are differences due to limitations of the ILI data. For example, holiday weeks result in biased estimates for percent of patient visits with ILI, as there tend to be fewer non-ILI patient visits in holiday weeks.

Figure 1

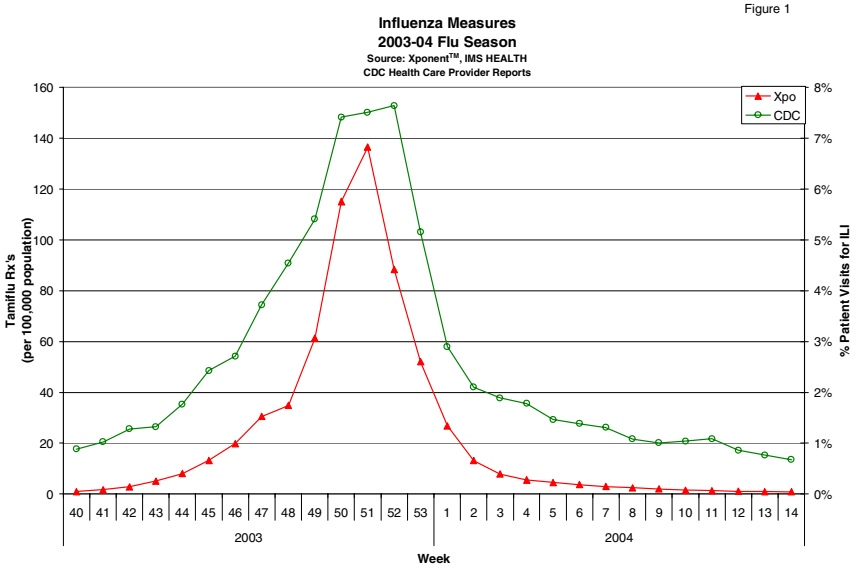
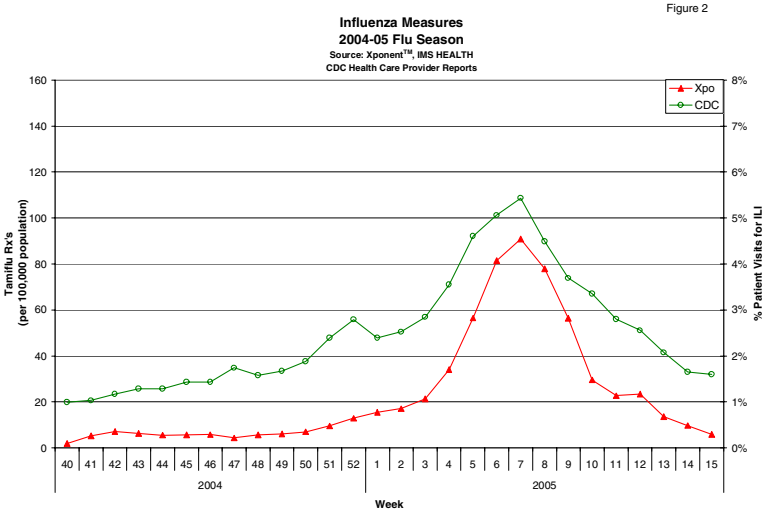


Figure 2



Analyzing the prescription data using CuSUM reveals the trend changes in influenza (Figures 3 and 4). Use of a Poisson model for the CuSUM allows examination of the length of the influenza season, and further highlights the differences in the two seasons. The 2003-2004 influenza season had run its course by the time the 2004-2005 influenza season had begun. In addition, the 2003-2004 influenza season achieved a magnitude roughly 30% above that of the 2004-2005 influenza season.

Use of a linear trend model allows examination of changes in the strength and direction of the outbreak. Here again, differences in the two outbreaks can be observed. The 2003-2004 influenza season showed both a sharper outbreak and ebb than did the 2004-2005 influenza season.

Figure 3

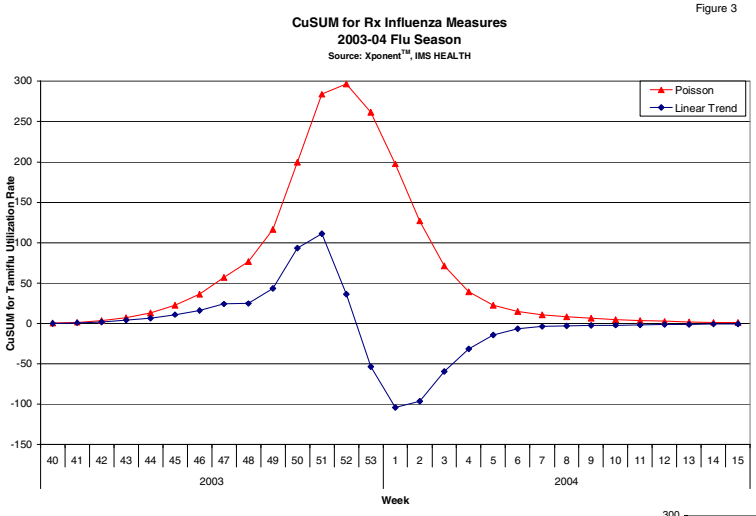
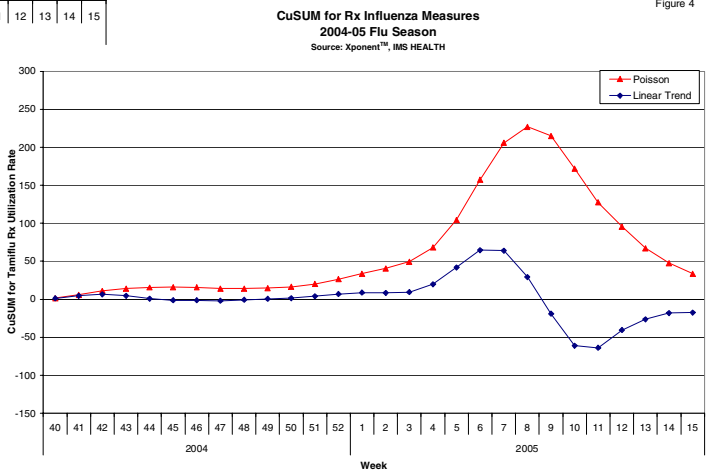


Figure 4

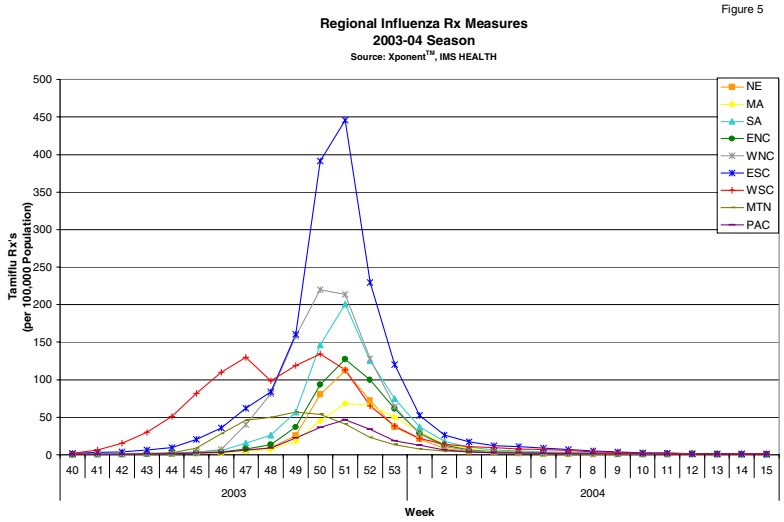


5.2 Regional Level

Examination of influenza measures by nine Census regions focused on the 2003-2004 influenza season. Looking at the prescription data (Figure 5), regional differences in terms of timing and magnitude of the outbreak can be observed.

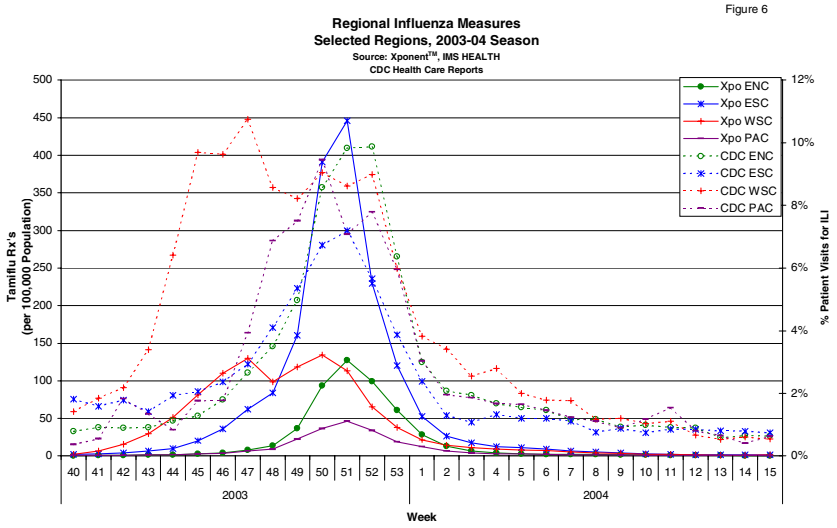
The West South Central region showed evidence of experiencing an earlier outbreak than in other regions, while the East South Central showed evidence of a more severe outbreak than occurred in the remaining regions.

Figure 5



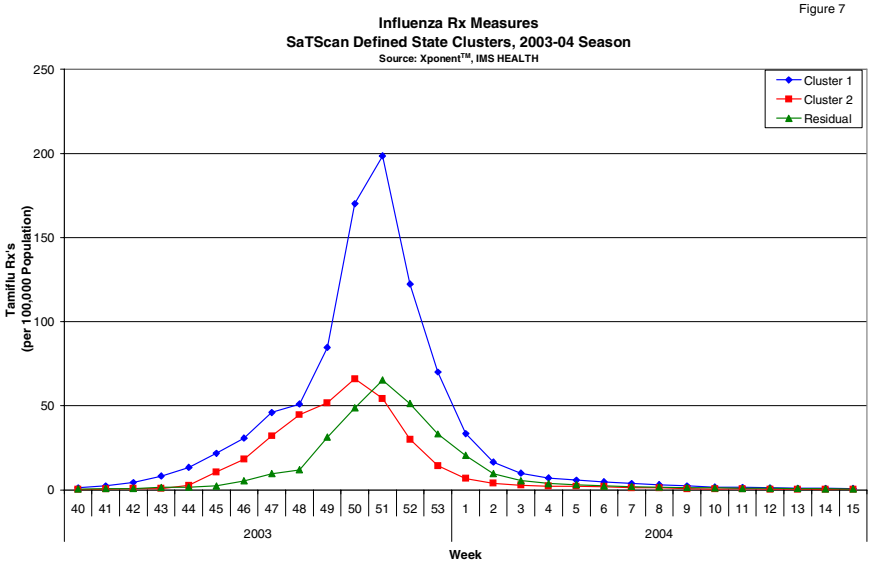
Comparison of the prescription data with the ILI data (Figure 6) indicates the timings of the outbreaks by region are consistent with those reported by the CDC, although the relative levels do appear to differ somewhat. This is likely due in large part to the differences in what is being measured between the two series.

Figure 6



The SaTScan software was used to geographically cluster the state level data on the basis of commonality in outbreak distributions. The results (Figure 7) showed three clusters of states. Cluster one, which included both the West South Central and East South Central, as well as other states, experienced an earlier and more severe outbreak than the other two clusters. The second cluster, consisting primarily of states from the West North Central region, had an earlier outbreak than the third cluster, although the severity for those two clusters were similar.

Figure 7



5.3 State Level

Prescription data for the East South Central and West South Central regions were examined at the state level (Figures 8 and 9). Within the East South Central region, Mississippi experienced the outbreak of influenza several weeks earlier than the other states in the region, while Kentucky experienced the outbreak several weeks later. All states within the region, however, experienced very similar severity of influenza outbreaks.

Within the West South Central region, more variability in state outbreak patterns was seen. Louisiana and Texas experienced earlier outbreaks than the remainder of the states within the region, while Arkansas experienced a more severe outbreak than the remainder of the states within the region.

Figure 8

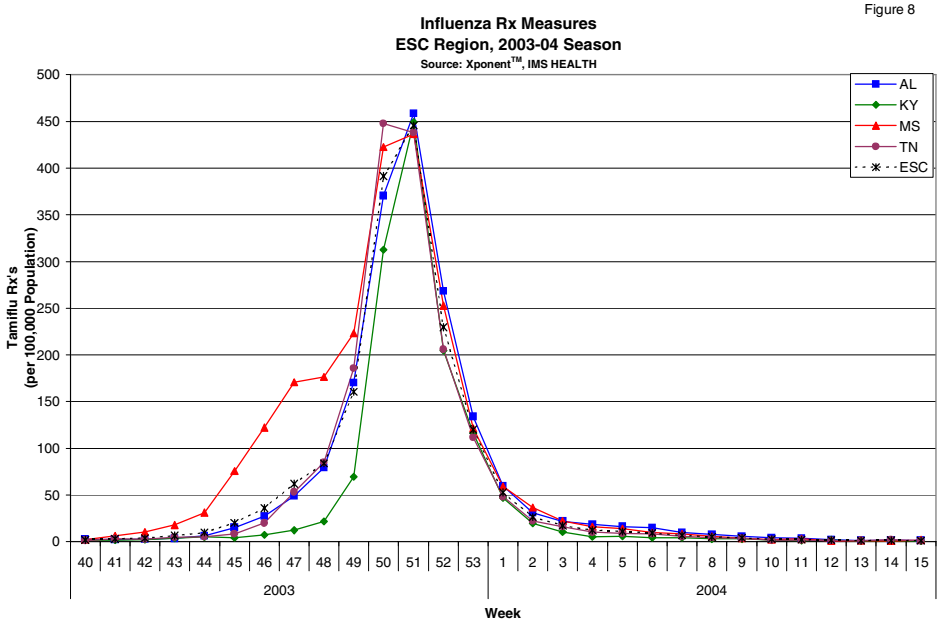
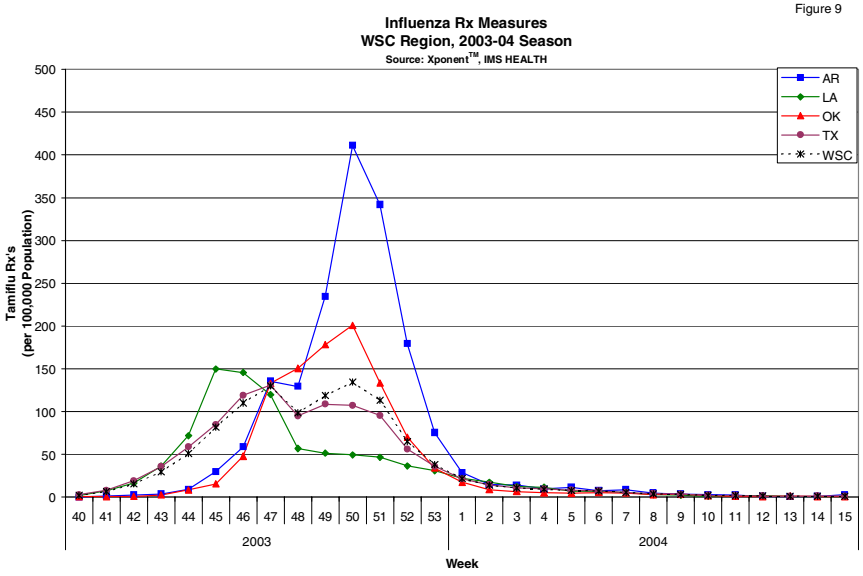


Figure 9



6. Summary

Prescription data for Tamiflu appear to reflect patterns of influenza as reported by CDC, across time both nationally and regionally. Relative severity of influenza outbreak can deviate between prescription and ILI data. However, these deviations may be related to differences between population and visit-based measures rather than to different signals from the two sources.

Thus, it appears the prescription data can provide useful information for influenza outbreak detection. The prescription data are timely and enhance capabilities for quantification and localization of outbreak detection, thereby addressing the initial detection, quantification, and localization factors for successful early detection.

7. Future Research

One near-term objective for future research is to seek to correlate the measures, prescriptions per 100,000 population, and percent patient visits for ILI, to understand reasons for differences in severity.

Improved detection will involve examining performance of detection models in identifying disease incidence. This would include the ability to distinguish among areas and seasonal patterns, such as the use of the SMART approach and/or the incorporation of error measures for prescription estimates. Further, more timely detection could be enabled through the use of daily prescription information, which would require development of small area estimation models.

Extension of the use of prescription data for detection requires correlation of the prescription usage with disease for other conditions, preferably chronic as well as acute diseases. Application of demographic data available for prescription data could also be considered for use in the investigation aspects of detection.

References

CDC (2005a). "Overview of Influenza Surveillance in the United States" http://www.cdc.gov/flu/weekly/pdf/flu-surveillance-overview.pdf.

CDC (2005b). *2003-2004 Influenza Season Summary.* URL: <http://www.cdc.gov/ncidod/diseases/flu/weeklychoice.htm>. Date accessed: July 29, 2005.

CDC (2005c). *2004-2005 Influenza Season Summary.* URL: <http://www.cdc.gov/ncidod/diseases/flu/weeklychoice.htm>. Date accessed: July 29, 2005.

Kulldorff, M. and Information Management Services, Inc. (2005) SaTScan™ v6.0: Software for the Spatial and Space-Time Scan Statistics. http://www.satscan.org.

Raubertas R.F. (1989) "An analysis of disease surveillance data that uses the geographic locations of the reporting units" Stat Med Vol 8, pp. 267-271.

Rolka, H. (2004) "Overview of Analytic Components of Early Detection" Public Health Information Network Conference.

Stoto, M.A., Schonlau, M., Mariano, L.T., (2004) "Syndromic Surveillance: Is It Worth the Effort?" Chance, Vol 17, No. 1, pp. 19-24.

U.S. Census Bureau (2005), Annual Population Estimates: States, http://www.census.gov/popest/states/NST-ann-est.html. Date accessed: July 28, 2005.