

Value of ICD-9–Coded Chief Complaints for Detection of Epidemics

Fu-Chiang Tsui¹ PhD, Michael M. Wagner¹ MD, PhD, Virginia Dato² MD, MPH,
Chung-Chou Ho Chang² PhD

¹Center for Biomedical Informatics and ²Graduate School of Public Health
University of Pittsburgh

To assess the value of ICD-9–coded chief complaints for early detection of epidemics, we measured sensitivity, positive predictive value, and timeliness of Influenza detection using a respiratory set (RS) of ICD-9 codes and an Influenza set (IS). We also measured inherent timeliness of these data using the cross-correlation function.

We found that, for a one-year period, the detectors had sensitivity of 100% (1/1 epidemic) and positive predictive values of 50% (1/2) for RS and 25% (1/4) for IS. The timeliness of detection using ICD-9–coded chief complaints was one week earlier than the detection using Pneumonia and Influenza deaths (the gold standard). The inherent timeliness of ICD-9 data measured by the cross-correlation function was two weeks earlier than the gold standard.

INTRODUCTION

Because of the threat of bioterrorism, improvement in the nation's capability to detect epidemics is of increasing importance^{1,2}. In particular, more timely detection is needed to potentially mitigate mortality, morbidity and economic costs³.

Timely detection of epidemics requires sources of data that themselves are timely; that is, types of data that will reflect the early "symptoms" of an epidemic and that are available in real-time, without delays due to manual data collection, transcription, or batch mode processing.

Thus, there is current interest in the use of existing sources of data— school or industrial absenteeism, grocery stores purchases, pharmacy purchases, web queries, and data collected during emergency room visits—for the detection of epidemics. Such sources of data form the basis of new detection systems being developed in New York City, Washington DC, and Pittsburgh⁴⁻⁶.

Surprisingly, relatively little work has been done to characterize the performance characteristics of detection systems that are based on such data. One study⁷ measured sensitivity, specificity, and timeliness of Influenza epidemic detector from health service data such as emergency home visits and drug sales.

There are at least two general questions about detection from routinely collected data: First, "What are the relative values of different types of routinely collected data for early detection?" and second, "What

types of algorithmic approaches are effective for analyzing these data?"

In this paper, we conduct research of the first type. We measure the value of ICD-9–coded chief complaints obtained routinely at the time of presentation to an Emergency Department (ED). Such data are of interest because they are collected routinely, are often available electronically in real-time, and—unlike other data used to monitor disease outbreaks (e.g., pneumonia incidence, or pneumonia deaths)—have the potential to detect sick individuals during early prodromal phases (i.e., while they are experiencing early, nonspecific symptoms).

In general, we cannot measure the value of data for detection directly. Instead, we must measure the detection performance achievable with the data by using the data as input to a detection system, and then measuring the performance of that detection system on one or more epidemic diseases. In this research, we built a detection system based on a standard detection algorithm (the Serfling Method) and measured its performance for Influenza. We selected Influenza because it is an excellent surrogate disease for research on detection systems for bioterrorism. Most militarily important biologic agents present with early clinical symptoms that resemble viral flu syndromes¹.

From measurements of timeliness, sensitivity, and positive predictive value of this detection system, we hope to gain insights about the potential of this type of data for epidemic detection.

METHODS

ICD-9–Coded Chief Complaints

ICD-9 is a diagnostic coding system that provides codes for chief complaints and diagnoses⁸. In our ED, chief complaints are coded at the time of presentation using ICD-9 codes. On arrival, the patient or a family member is interviewed by a triage nurse who writes the chief complaint in free text on a paper form. A registration clerk then enters the complaint using ICD-9 in the registration system.

Study Period

We used standard epidemiological numbering in which we refer to week 1 of 1999 as "1/99" and week 52 as "52/99". As defined in the mortality and morbidity weekly reports (MMWR) Table IV⁹, each week starts from Sunday instead of the date that January 1 starts as described in¹⁰. The study period was the 52-week

period from 49/99 (December 4, 1999) through 48/00 (December 1, 2000).

“Respiratory” and “Influenza” ICD-9 Sets

In this research, we evaluated two different sets of ICD-9 codes as the basis of our detection scheme: a “respiratory” set (RS) and an “Influenza” set (IS).

The RS is motivated by the observation that most bioterrorism diseases present early with respiratory symptoms. The IS is motivated by the fact that we are using Influenza as a surrogate disease in this research and therefore we want to build as good a detector as possible for that disease from the available ICD-9-coded chief complaint data.

We developed these ICD-9 sets as follows: In earlier research described in ¹¹, two internists created RS and a “viral” set. In particular, the internist reviewed all ICD-9 codes that had been used to represent chief complaints in the ED during the past three years. The goal of that project was to create *sensitive* detection systems, so the internists erred on the side of including too many ICD-9 codes. They assigned 64 codes to RS and 33 to the “viral” set.

We used the RS without modification in this research. It included disease codes and symptom codes for entities such as pneumonia, asthma, cough, and dyspnea.

To create IS, one of the internists edited the combined RS and “viral” set. The rationale for combining these sets was that Influenza presents with respiratory and viral-like features. To create a detector for influenza, he excluded ICD-9 codes from the combined set that he thought would reduce specificity without affecting sensitivity. The resulting IS contained 35 codes.

The Serfling Method

The Serfling Method ¹² is a methodology for automatic detection of epidemics from time-series data (e.g., our weekly ED visit counts). We used it as the method for detecting epidemics. The Serfling Method is a cyclical linear regression model that uses a linear trend and two harmonic terms to create a model of non-epidemic periods. A threshold curve based on statistical expectations is established, e.g., 95% confidence interval of the baseline. A Serfling-based detection system will signal an epidemic whenever the observed time-series data exceed the threshold. This method has been applied to nationwide pneumonia and influenza (P&I) deaths in the USA and influenza-like-illness in France ¹³⁻¹⁵.

The equation for the Serfling Method is,

$$y(t) = a + b * t + c * \sin(2\pi \frac{t}{52}) + d * \cos(2\pi \frac{t}{52}) \quad (1)$$

where $y(t)$ is the prediction of the model at week t , assuming no epidemic is ongoing, and parameters a , b , c , and d are model coefficients. When used for Influenza modeling, the two harmonic terms are set to

period of 52 weeks to account for seasonal effect. Minor justification for the period (52 weeks) of the two harmonic terms was addressed in ¹⁴ when leap years were considered. The data collection period for RS and IS was from year 1997 to year 2000 that were non-leap years.

Establishing the Serfling Baseline

The Serfling model requires training data (e.g., weekly counts of ED visits for RS or IS) from non-epidemic years to parameterize a base model. We had data available for the years 1998 and 1999. Data for Influenza from consecutive, non-epidemic years is difficult to obtain because of the regularity of Influenza, and 1998 and 1999 were epidemic years. A solution is to manually remove counts during epidemic weeks from the training data, but finding documented epidemic weeks is also difficult.

We were unable to find a standard automatic procedure for removing epidemic weeks; therefore we used the following five-step procedure to remove data points that could possibly fall into epidemic weeks. The assumption underlying this procedure is that the outliers above the 95% confidence interval that we are removing are data occurring within epidemic weeks.

1. Obtain coefficients for the regression model by fitting the raw training data without any manipulation.
2. Compute the preliminary baseline based on the coefficients obtained in Step 1.
3. Remove outliers that are above one-sided 95% confidence interval (CI) of the baseline obtained in Step 2. Here we cut off the outliers with considering the seasonal effect, the preliminary baseline.
4. Refit the model to the data without outliers and recalculate the model coefficients.
5. Use the model to predict next year’s baseline curve.

We used the first 100 weeks of RS and IS counts (week 1/98 to week 48/99) as training data (after removing the outliers using the above five steps).

The Detection System

The detection system in this study used the above Serfling Method, set to trigger an alarm whenever the time-series weekly counts (i.e., RS counts or IS counts, depending on which was being studied) exceeded threshold levels for two or more consecutive weeks. The threshold levels were defined to be the one-sided upper 95% limit of the expected baseline. This approach follows standard practice ^{13, 14}.

We used SAS[®], a commercial statistical tool, to implement the Serfling model.

The Gold Standard

A gold standard is a tool used to evaluate a system or study, implying general acceptance or consensus about its validity as a measure of a system ¹⁶. Epidemic periods defined on the basis of pneumonia and

influenza (P&I) deaths, often times using the Serfling Method, are used in public health practice and research as the primary quantitative detection index for influenza¹²⁻¹⁴.

To identify the epidemic periods in Pittsburgh, we applied the Serfling Method to P&I death data for the three-year period 2/97* to 48/99 for training and then used the calibrated Serfling model to identify epidemic occurrences during the 52-week study period (weeks 49/99 to 48/00). We used the standard definition of the onset of an influenza epidemic, which is taken to be the beginning of any two-week (or longer) period during which P&I deaths exceeded predicted levels¹⁴. We also used the standard definition of the end of an influenza epidemic to be the beginning of the first two-week period during which P&I deaths returned to non-epidemic levels as defined by the Serfling Method¹⁴.

We obtained P&I death data for Pittsburgh for this period from CDC's MMWR Table IV⁹. The weekly P&I deaths for Pittsburgh on the MMWR Table IV were incomplete due to delays in issuing of death certificates, so we augmented this data with the reports from the Allegheny County Health Department.

Measurement of Sensitivity, Positive Predictive Value (PPV), and Timeliness

For detection systems based on RS and IS, we calculated sensitivity, PPV, and timeliness.

We defined a true positive alert to be when the Serfling Method signaled the onset of an epidemic within four weeks before or two weeks after the P&I-death defined epidemic (a standard methodological approach)¹²⁻¹⁴. We defined sensitivity as the ratio of true positive epidemic alerts given by RS (or IS) detection systems to the "gold standard" alerts given by the P&I detection system. We defined PPV as the ratio of true positive epidemic alerts to the total alerts given by the RS (or IS) detection system. We calculated standard epidemiological measure of timeliness by subtracting detection dates computed from P&I data by the Serfling Method from those computed from RS and IS.

The Cross-Correlation Function

We also measured the inherent timeliness of the data themselves using the cross-correlation function. The cross-correlation function is a well-established tool used in the field of signal processing for estimating times lags between time series¹⁷. For example, if we have time series whose peaks are in temporal alignment, the maximum value of the cross-correlation function will occur at a time lag of zero.

The assumption underlying use of the cross-correlation function is that the time series are stationary. Thus, before employing the cross-

correlation function, we removed the mean and linear trend of the ICD-9 counts and the P&I deaths.

The following equation is the estimate of the cross-correlation function $r_{xy}(m)$, a second-order statistics, between time series x and y ,

$$r_{xy}(m) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) y(n+m)$$

where m ranges from 0 to $N-1$.

The ED registration records included a time stamp, which we use in this study as the time when the data were available for the purpose of disease surveillance. In the P&I death data, the time we used was the time of death.

We computed cross correlations between RS and P&I deaths and between IS and P&I deaths for two years (weeks 49/98-48/00).

RESULTS

Figures 1-3 show the weekly counts, baselines, and thresholds from 1998 to 2000 for RS, IS, and the P&I gold standard data.

The following equations for $y_r(t)$ and $y_f(t)$ are the parameterized Serfling baselines for the RS and IS data, respectively, where t is the week.

$$y_r(t) = 30.66 + 0.083t + 1.50 \sin(2\pi \frac{t}{52}) + 6.90 \cos(2\pi \frac{t}{52})$$

$$y_f(t) = 65.34 + 0.271t + 3.45 \sin(2\pi \frac{t}{52}) + 8.41 \cos(2\pi \frac{t}{52})$$

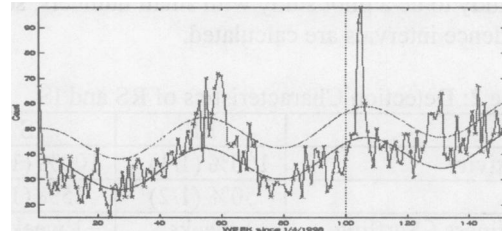


Figure 1: Weekly counts of RS from 1/98 to 48/00. We used the first 100-week period for training in the Serfling Method. We used the remaining 52 weeks, separated by the vertical dotted line, for evaluation.

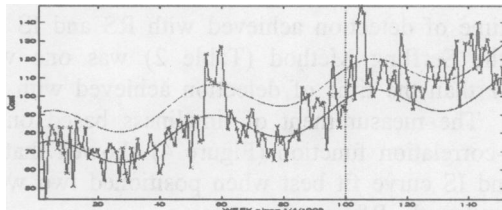


Figure 2: Weekly counts of IS from week 1/98 to 48/00.

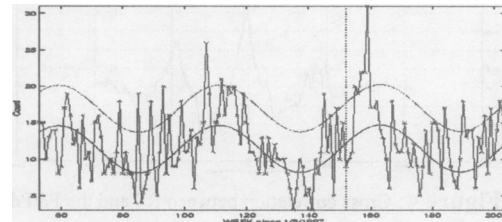


Figure 3: Weekly counts of P&I from week 1/98 to 48/00.

*Since Year 1997 had total 53 weeks, we started from 2/97 (Jan. 5, 1997) to maintain the 52-week period in the harmonic terms.

DISCUSSION

Table 1 gives the start and end weeks of all epidemic periods—those identified by the gold standard (the Serfling Method applied to Pittsburgh P&I data), and the RS- and IS-based detectors. Using P&I or RS data, the detection system identified three epidemic periods. Using IS data, it identified four. In Table 1, the epidemics detected by the three detectors are placed in the same row if the start date of the non gold standard methods fell within the six-week window (four-weeks before and two after) of the P&I defined epidemics. Table 1 also includes the timeliness (in parentheses) calculated by subtraction of the date of the P&I defined epidemic from the date of the ICD-9 defined epidemic.

Table 1: Epidemics Detected by P&I, RS, and IS

P&I Deaths (gold standard)	RS	IS
1/00 – 3/00	52/99 – 2/00 (+1 week*)	52/99 – 2/00 (+1 week)
None	None	14/00 – 15/00
None	23/00 – 27/00	23/00 – 32/00
None	None	35/00 – 43/00

* Timeliness is measured in weeks as described in the text. A positive value represents an improvement relative to P&I deaths.

Table 2 shows the sensitivity and PPV for RS and IS. Due to the small number of epidemics, we considered this study to be a pilot study with small numbers, so no confidence intervals are calculated.

Table 2: Detection Characteristics of RS and IS

	RS	IS
Sensitivity	100% (1/1)	100% (1/1)
PPV	50% (1/2)	25% (1/4)
Timeliness (Serfling Method)	+1 weeks	+1 weeks
Timeliness (cross-correlation function)	+ 2 weeks	+ 2 weeks

The time of detection achieved with RS and IS data and the Serfling Method (Table 2) was one week *earlier* than the time of detection achieved with P&I data. The measurement of timeliness based on the cross-correlation function (Figure 4) showed that the RS and IS curve fit best when positioned two weeks *earlier* than the P&I curve.

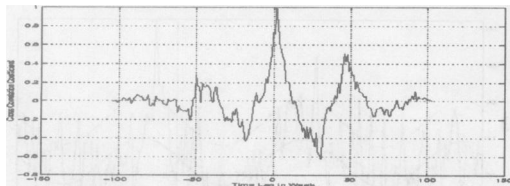


Figure 4: Cross correlation between RS and the P&I deaths.

Due to the singular Influenza epidemic in the study period, our measurements of sensitivity, PPV, and timeliness need to be confirmed by a larger study. The results nevertheless suggest that the sensitivity and timeliness achievable with ED ICD-9-coded chief complaints can be good. The PPVs of the two ICD-9 signals, RS and IS, were lower. The false positive might be partly explained by the nonspecific characteristic of ICD-9-coded chief complaints being caused possibly by any respiratory outbreak such as asthma or viral illnesses, whereas the P&I death signal was more specific to Influenza.

As we would expect, ED visits for respiratory and Influenza symptoms precede deaths from pneumonia and Influenza.

The Serfling Method

The Serfling method is a static regression model with parameters, *e.g.*, *a*, *b*, *c*, and *d* in Eq. (1), that do not vary with time after they are optimized through the training data. When applied to less noisy data, the model usually predicts baseline well during the study (prediction) period. Thus, fewer false positive alerts based on the 95% CI of the baseline would be expected. In this study, P&I data seemed to be less noisy, probably because it was specific to Influenza and represented the complete surveillance results for Pittsburgh. As shown in Fig. 3, the predicted baseline of P&I data matched the observed data during the study period. However, the predicted baselines for RS and IS seemed to deviate from the observed data, as shown in Figures (1) and (2). This deviation could indicate that RS and IS were noisy due to their nonspecific characteristics and limited surveillance scope to a single ED. The implication of this observation for future research is that either (1) alternative detection algorithms should be used, or (2) if the Serfling Method is used, the dependence of the method on signal properties must be respected.

This study was based on an assumption that the Serfling Method was a reasonably good method for detection, and therefore the detection performance achievable using that method on ICD-9-coded chief complaints would be informative about that data source. We selected the Serfling Method because it is a standard method in epidemiological research. Further conclusions about the value of ICD-9-coded chief complaints for early detection of epidemics should be based on results using additional detection algorithms—such as autoregressive integrated moving average (ARIMA), adaptive filters, and hidden Markov model (HMM) methods.

Influenza

The study also rests on the assumption that the value of ICD-9-coded chief complaints for detection of Influenza will be informative about their potential for

other epidemics such as bio-terrorist release of aerosol Anthrax. This belief is probably justified because Influenza would be harder to detect having lower incidence and broader time distribution of cases. Influenza probably represents a more difficult detection problem than does a large-scale bio-terroristic release.

The Value of ICD-9-coded Chief Complaints

The observed capability to detect Influenza epidemics using ICD-9-coded chief complaints, the Serfling Method, and data from a single ED are probably useful for public health surveillance purposes.

From our previous discussion about signal properties, the PPV might be improved by increasing the scope of the surveillance system to include larger populations, by utilizing different detection algorithms, or by using more complete ICD-9 codes for respiratory syndrome (if we want to put the detector in a different hospital).

In this study, we used weekly counts. By monitoring daily or even hourly counts of ICD-9-coded chief complaints, the timeliness of an epidemic alert can be further improved.

The use of the cross-correlation function to measure the relative timeliness of different types of data for early detection is a new application of an existing technique. We believe it will be useful for screening potential sources of data for inclusion in a detection scheme. In this study, the cross-correlation function demonstrates that the signal provided by ICD-9-coded chief complaints occurs two weeks earlier than the signal provided by P&I data.

Our belief that ICD-9-coded chief complaints can be used for early detection is also based on our results¹¹ showing good sensitivity (0.44) and specificity (0.97) for detection of cases of acute respiratory illness using the ICD-9-coded chief complaints.

These methods and results may apply to diseases that present with non respiratory and non viral prodromes. We and others have also developed but not tested diarrheal, encephalitic, botulinic, rash, viral, and hemorrhagic sets of ICD-9 codes.

ACKNOWLEDGEMENT

We thank Drs. Jeremy Espino, Ted Tsai, Andrew Post, and Janine Janosky. We also thank Dr. D. Hennon at Allegheny County Department of Health for Pittsburgh P&I data. This work was supported by grants GO8 LM06625-01, and T15 LM/DE07059 from the National Library of Medicine; contract 290-00-0009 from the Agency for Healthcare Research and Quality. This article was also supported by Cooperative Agreement Number U90/CCU318753-01 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.

REFERENCES

1. Institute of Medicine National Research Council. *Chemical and Biological Terrorism*. Washington: National Academy Press, 1999.
2. Henderson, D. A. "Bioterrorism as a Public Health Threat," *Emerging Infectious Diseases* 4 (1998).
3. Kaufmann, A., Meltzer, M., and Schmid, G. "The Economic Impact of a Bioterrorist Attack: are prevention and post attack intervention programs justifiable," *Emerging Infectious Diseases* 3 (1997): 83-94.
4. Pavlin, J. "Electronic surveillance system for the early notification of community-based epidemics (ESSENCE)," *Conference and workshop on syndromic and other surveillance methods for emerging infections including bioterrorism*. Gaithersburg, MD, 2000.
5. Trent, L. "Global Expeditionary Medical System (GEMS)," *Conference and Workshop on Syndromic and Other Surveillance Methods for Emerging Infectious Including Bioterrorism*. Gaithersburg, MD, 2000.
6. RODS. <http://ultra.cbmi.upmc.edu/~jue/rods/>
7. Quenel, P., Dab, W., Hannoun, C., and Cohen, J. M. "Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics," *International Journal of Epidemiology* 23 (1994): 849-55.
8. ICD-9. "Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death. Based on Recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-Ninth World Health Assembly." Geneva, Switzerland, 1977.
9. MMWR Table IV. <http://wonder.cdc.gov/mmwr/mmwrmort.asp>
10. Hutwagner, L. C., Maloney, E. K., Bean, N. H., Slutsker, L., and Martin, S. M. "Using Laboratory-Based Surveillance Data for Prevention: An Algorithm for Detecting Salmonella Outbreaks," *Emerging Infectious Diseases* 3 (1997): 395-400.
11. Espino, J., and Wagner, M. M. "Accuracy of ICD-9-coded Chief Complaints and Diagnoses for the Detection of Acute Respiratory Illness," *AMIA 2001 Annual Symposium*. Washington, DC, 2001.
12. Serfling, R. E. "Methods for Current Statistical Analysis of Excess Pneumonia-Influenza Deaths," *Public Health Reports* 78 (1963): 494-506.
13. Lui, K.-J., and Kendal, A. "Impact of Influenza Epidemics on Mortality in the United States from October 1972 to May 1985," *American Journal of Public Health* 77 (1987): 712-716.
14. Simonsen, L., Clarke, M. J., Williamson, G. D., Stroup, D. F., Arden, N. H., and Schonberger, L. "The Impact of Influenza Epidemics on Mortality: Introducing a Severity Index," *American Journal of Public Health* 87 (1997): 1944-1950.
15. Tobiana, L., and Flahault, A. "A Space-Time Criterion for Early Detection of Epidemics of Influenza-like-illness," *European Journal of Epidemiology* 14 (1998): 465-470.
16. Friedman, C. P., and Wyatt, J. C. *Evaluation Methods in Medical Informatics*: Springer-Verlag, 1997.
17. Shumway, R. H., and Stoffer, D. S. *Time Series Analysis and Its Applications*. New Yprk: Springer, 2000.