Research article

Time series modeling for syndromic surveillance

Ben Y Reis*[1] and Kenneth D Mandl[2]

Address: [1]Children's Hospital Informatics Program, Boston, Massachusetts and [2]Harvard Medical School, Boston, Massachusetts

Email: Ben Y Reis* - reis@mit.edu; Kenneth D Mandl - kenneth_mandl@harvard.edu

* Corresponding author

Abstract

Background: Emergency department (ED) based syndromic surveillance systems identify abnormally high visit rates that may be an early signal of a bioterrorist attack. For example, an anthrax outbreak might first be detectable as an unusual increase in the number of patients reporting to the ED with respiratory symptoms. Reliably identifying these abnormal visit patterns requires a good understanding of the normal patterns of healthcare usage. Unfortunately, systematic methods for determining the expected number of (ED) visits on a particular day have not yet been well established. We present here a generalized methodology for developing models of expected ED visit rates.

Methods: Using time-series methods, we developed robust models of ED utilization for the purpose of defining expected visit rates. The models were based on nearly a decade of historical data at a major metropolitan academic, tertiary care pediatric emergency department. The historical data were fit using trimmed-mean seasonal models, and additional models were fit with autoregressive integrated moving average (ARIMA) residuals to account for recent trends in the data. The detection capabilities of the model were tested with simulated outbreaks.

Results: Models were built both for overall visits and for respiratory-related visits, classified according to the chief complaint recorded at the beginning of each visit. The mean absolute percentage error of the ARIMA models was 9.37% for overall visits and 27.54% for respiratory visits. A simple detection system based on the ARIMA model of overall visits was able to detect 7-day-long simulated outbreaks of 30 visits per day with 100% sensitivity and 97% specificity. Sensitivity decreased with outbreak size, dropping to 94% for outbreaks of 20 visits per day, and 57% for 10 visits per day, all while maintaining a 97% benchmark specificity.

Conclusions: Time series methods applied to historical ED utilization data are an important tool for syndromic surveillance. Accurate forecasting of emergency department total utilization as well as the rates of particular syndromes is possible. The multiple models in the system account for both long-term and recent trends, and an integrated alarms strategy combining these two perspectives may provide a more complete picture to public health authorities. The systematic methodology described here can be generalized to other healthcare settings to develop automated surveillance systems capable of detecting anomalies in disease patterns and healthcare utilization.

Background

The earliest detectable sign of a covert germ warfare attack may be unusual increases in the number of people seeking healthcare. For example, victims of an anthrax attack

might present to emergency departments with flu-like or respiratory syndromes [1]. The identification of abnormally high visit rates relies on a good understanding of the normal patterns of healthcare utilization. Once expected rates of utilization are established, alarm thresholds can be set for public health alerts.

Most previous models of healthcare usage have not incorporated real time data and were generally built to improve resource management, rather than to detect diseases [2–4]. Recently, the growing bioterrorist threat has led to the development of real-time *syndromic surveillance* systems built on top of existing information systems [5–7], with increased research activity in the biomedical engineering [8], public health [9], and general scientific [10] communities.

A number of different approaches have been developed for syndromic surveillance, with systems monitoring over the counter drug sales [11], web-based physician-entered reports [12], consumer health hotline telephone calls [13], and ambulatory care visit records [14]. However, due to the timeliness and data availability considerations, the majority of systems to date have focused on monitoring visit data from emergency department information systems. Several regional syndromic surveillance systems that focus on real time clinical and administrative emergency department datasets have been deployed [15–17].

Two data elements consistently relied upon are the International Classification of Disease (ICD) diagnostic codes and ED chief complaints (CC). ICD codes are universally used in the United States for hospital billing. Many EDs also record information about patients' presenting chief complaints. ICD and CC data often become available in databases during or shortly after a patient's visit, thus providing a basis for real time surveillance. CCs vary in quality because they are recorded prior to physician involvement in care, but on the other hand are timelier because they are elicited and recorded upon arrival to the ED. The ICD billing codes tend to be more accurate, though the delay in their availability ranges from hours to weeks [18,19]. To enable interoperability between different systems, data interchange standards are being proposed for communicating emergency department data for public health uses including syndromic surveillance [20].

Our goal is to be able to detect a possible anthrax attack by detecting an anomalous surge in emergency department usage and particularly in patients presenting with respiratory and flu-like syndromes. Using CC data, we sought to define the expected daily rates for emergency department total volume as well as for the frequency of visits of patient with flu-like and respiratory illness. These expected rates can be compared with actual rates in attempting to identify anomalies.

In addition to presenting a generalized framework for biosurveillance model development, the paper tests the model with simulated outbreaks, and includes a proposal for handling specific alarm-related issues that arise with ARIMA-based models. Furthermore, the present study focuses on data from a pediatric population, something not prevalent in the prior literature.

Methods

The record review was approved by the Children's Hospital Boston Committee on Clinical Investigation (Protocol Number M-01-03-040). The setting was Children's Hospital Boston, an urban, tertiary care pediatric teaching facility. Subjects were consecutive patients visiting the emergency department from June 2, 1992 through January 4, 2002. Patients with respiratory illness were identified using a previously validated set of chief complaint and ICD codes.

We built models of ED utilization using a time-series analytic approach. Models were constructed through an iterative process and were trained on roughly the first eight years of data (2,775 days). The remaining two full years of data (730 days) were reserved for testing and validation. The model and the statistical methods used in its development are described below. Complete details are also available from the authors. Our approach consisted of three main phases: signal characterization, model development, and model validation.

Signal characterization

As a first step towards understanding the trends and behaviours of ED utilization, we performed a principal Fourier component analysis on the daily visit totals (Figure 1). A Fast Fourier Transform (FFT) was calculated from the time series data, and the results indicated the presence of strong weekly and yearly periodicities in the data. The analysis was performed using the Matlab software package, Release 12, Version 6.0 [21]. Guided by these findings, we calculated the ensemble average profiles for both the weekly and yearly periodicities (Figures 2 and 3). These profiles represent the average number of visits for each day of the week and year respectively.

Model development

Our model development methods followed established time series forecasting procedures. The signal was broken up into its various components: the overall mean, the weekly trend and the yearly trend. Mean Absolute Percentage Error (MAPE) was used to measure of quality of fit. A MAPE of 0% indicated a perfect fit of the model to the training dataset.
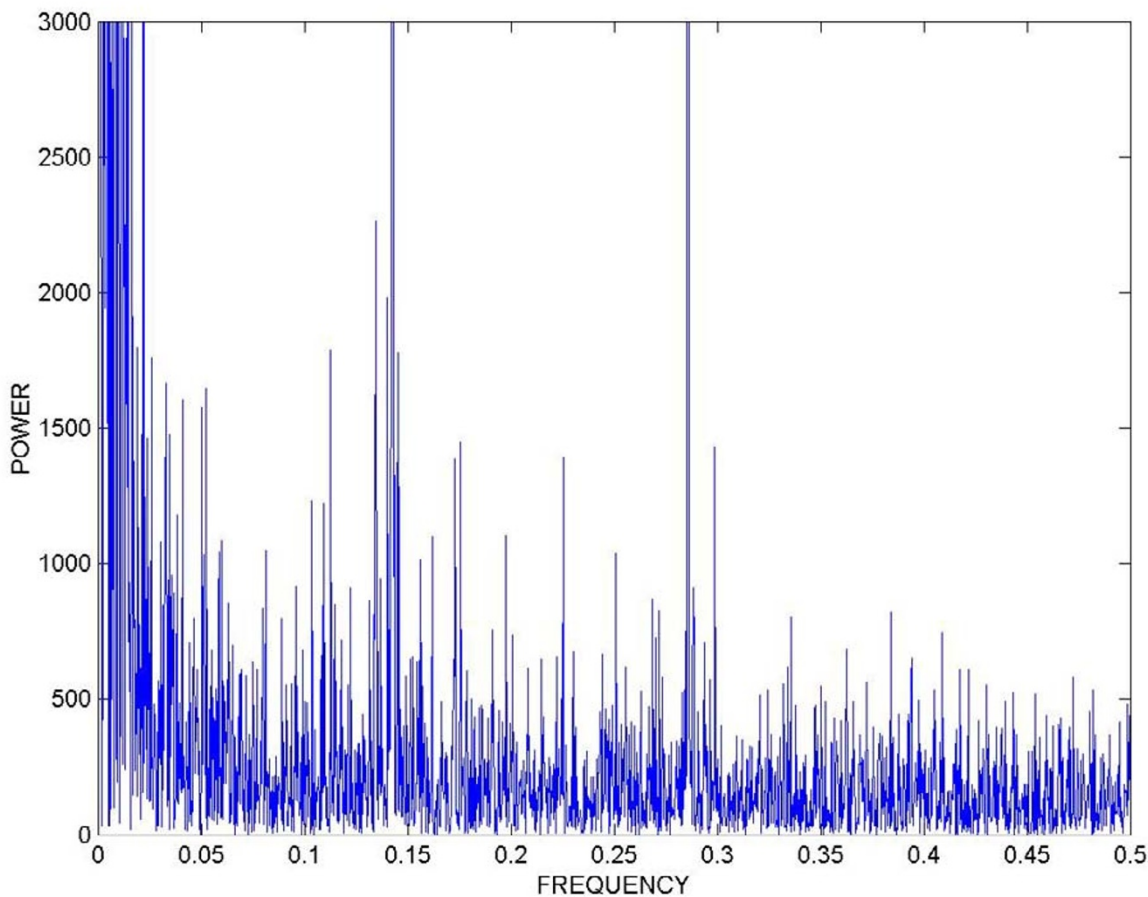
Figure 1
Fourier analysis. Fourier analysis of daily visit totals reveals strong weekly and yearly periodicities. Power is shown on the Y axis with frequency (1/day) shown on the X axis.

The trimmed mean seasonal model was then constructed from the overall mean, the weekly trend, and the yearly trend, as follows:

Expected number of visits = Overall mean.+ Mean for day of week + Trimmed mean for day of year

We started with a simple-mean model that consistently predicted 137 total visits every day, 48 of which were respiratory-related. This mean was subtracted from the data. Next the weekly average profile of the resulting signal was calculated, yielding a model of the average utilization patterns for each day of the week. The weekly profile was also subtracted out from the data. The yearly average profile of the resulting signal was calculated, yielding a model of the

average utilization patterns for each day of the year. To increase the robustness of this model and ensure that it was not overly sensitive to outliers in the training set, a trimmed mean function (ignoring a certain percent of the highest and lowest values) was used for calculating the yearly profiles. Testing a range of trimming magnitudes, we found that excluding the highest and lowest 25% of the values yielded the best results as far as improving the quality of fit (low MAPE).

When analyzing the residuals (forecasting errors) of the trimmed-mean seasonal model, a significant degree of autocorrelation was detected. Significant autocorrelations indicate that, for example, a high utilization rate on one day is associated with a high utilization rate on the
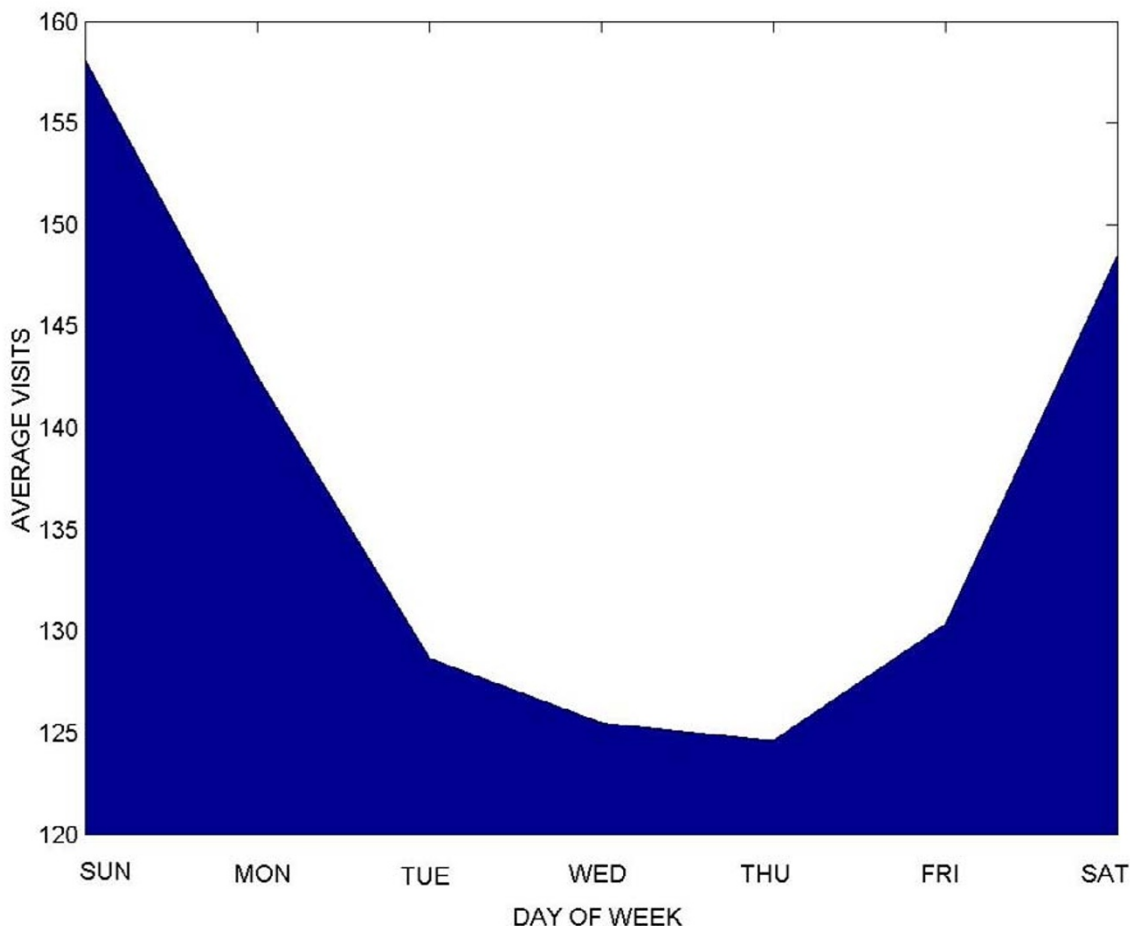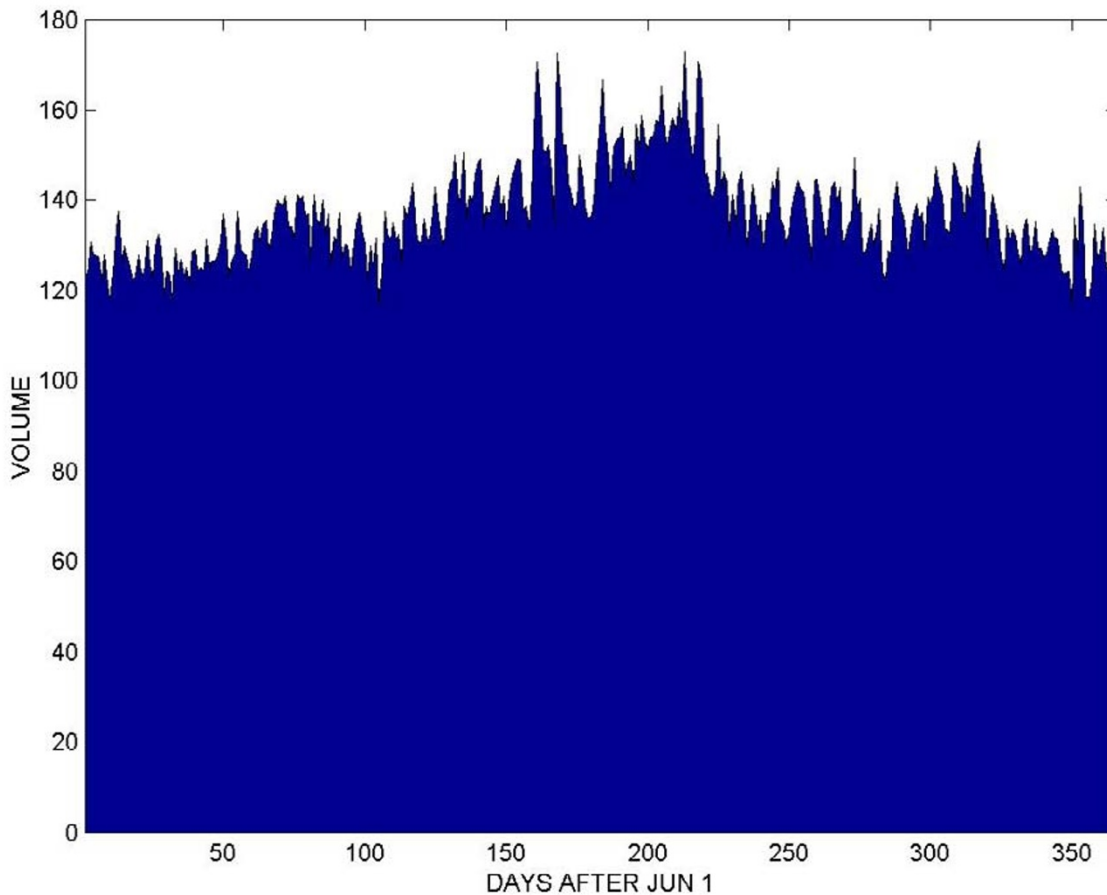
Figure 2
Weekly patterns. Weekly ensemble average of daily visit totals, shown from Sunday through Saturday, reveal a peak in visits on the weekends.

following day. To model these localized auto-correlative effects, we fit an Auto-Regressive Integrated Moving Average (ARIMA) time-series forecasting model to the residuals (forecasting errors) of the trimmed mean seasonal model. This class of models is often used in economic time series forecasting, and is described in full detail in [22]. The ARIMA model was fit using the SAS Release 8.2 software package Time Series Forecasting System (Cary, NC) [23].

Three main parameters need to be selected when fitting an ARIMA model: the order of autocorrelation (AR), the order of Integration (I), and the order of Moving Average (MA). The higher-order models are more complex and can usually achieve a better fit of the training data set, while the simpler low-order models are usually less likely to over-fit to training dataset.

The first derivative of the data was taken to ascertain whether Integration would improve the model. Since forecasting was not improved, the optimal degree of integration was determined to be zero (I = 0). Different combinations of AR and MA orders were then tested. Of all the models tested, an ARIMA(2,0,1) model (second-order Auto-Regressive combined with first-order Moving Average) was found to work best for overall ED volume. An ARIMA(1,0,1) model (first-order Auto-Regressive combined with first-order Moving Average) was found to work best for respiratory-related ED volume. The SAS toolkit

Figure 3
Yearly patterns, total visits Yearly ensemble average of daily visit totals, shown from June through June, reveals an increase in visits during the winter months.

was used to automatically fit the parameters for the models.

Model validation
The models, trained on approximately the first eight years of data – the training set, were validated using the final two years of the dataset – the validation dataset. Again, a MAPE of 0% indicated a perfect fit of the model to the validation dataset.

We set out to quantify the detection performance of a biosurveillance system based on the above ARIMA model of overall ED volume by measuring its ability to detect outbreaks. Since the historical dataset used here was devoid of any known outbreaks and there is a paucity of data

available on actual germ warfare attacks [11], we introduced a set of simulated outbreaks into the historical visit data by adding a certain number of simulated visits to specified days.

The simulated outbreaks lasted seven days each. While real outbreaks may last well beyond these durations, we focused on the first few days since useful detection systems should be able to spot outbreaks within that time-frame.

The outbreaks were spaced 15 days apart and there were 233 simulated 7-day outbreaks in total inserted into the data. This ensured that all the effects of any previous outbreak could be reset from the detection system's memory
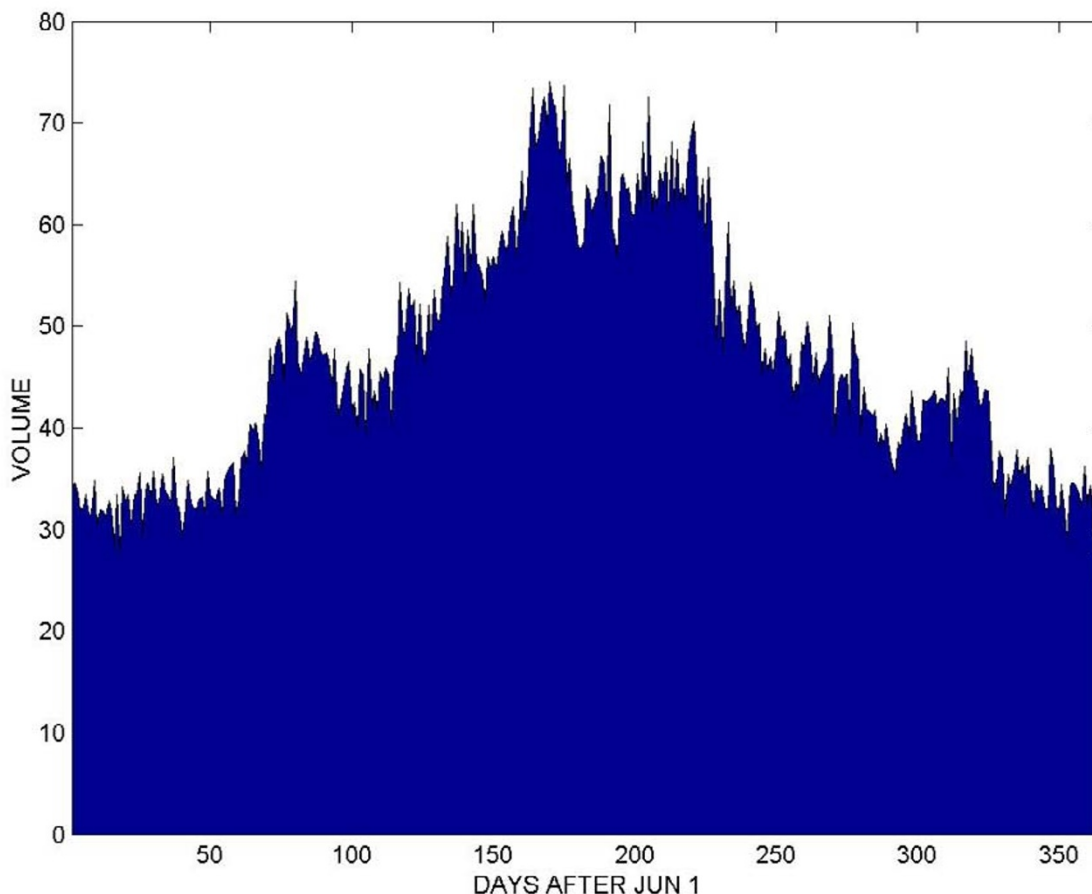
Figure 4
Yearly patterns, respiratory visits Yearly ensemble average of daily respiratory-related visit totals, shown from June through June, reveals peaks in visits during the fall, winter and spring.

before the onset of the next outbreak. Furthermore, since the only significant periodicities present in the data were seven day (weekly) and 365.25 day (yearly), the 15-day spacing of outbreaks yielded a good unbiased sample of days for infusing outbreaks.

Each complete 3,505-day simulation used outbreaks of only one size. Different sizes of outbreaks were tested in separate simulations. For each simulation, sensitivity was calculated as the ratio of the number of outbreaks detected (true positives) to the total number of outbreaks. Since there is an inherent tradeoff between sensitivity and specificity, a benchmark specificity of 97% was chosen, producing approximately one false alarm per month. A study

that evaluates different detection approaches using simulated outbreaks can be found in [24].

Results
Signal characterization
The data exhibited several periodicities. The weekly plot (Figure 2) revealed significantly higher ED utilization on the weekends, when primary care clinics tend to be closed. The yearly plot (Figure 3) showed an overall increase in utilization during the winter. A separate yearly ensemble average profile of respiratory-related visits (Figure 4) showed a general increase in respiratory-related visits during the winter, together with localized peaks in the fall and spring.
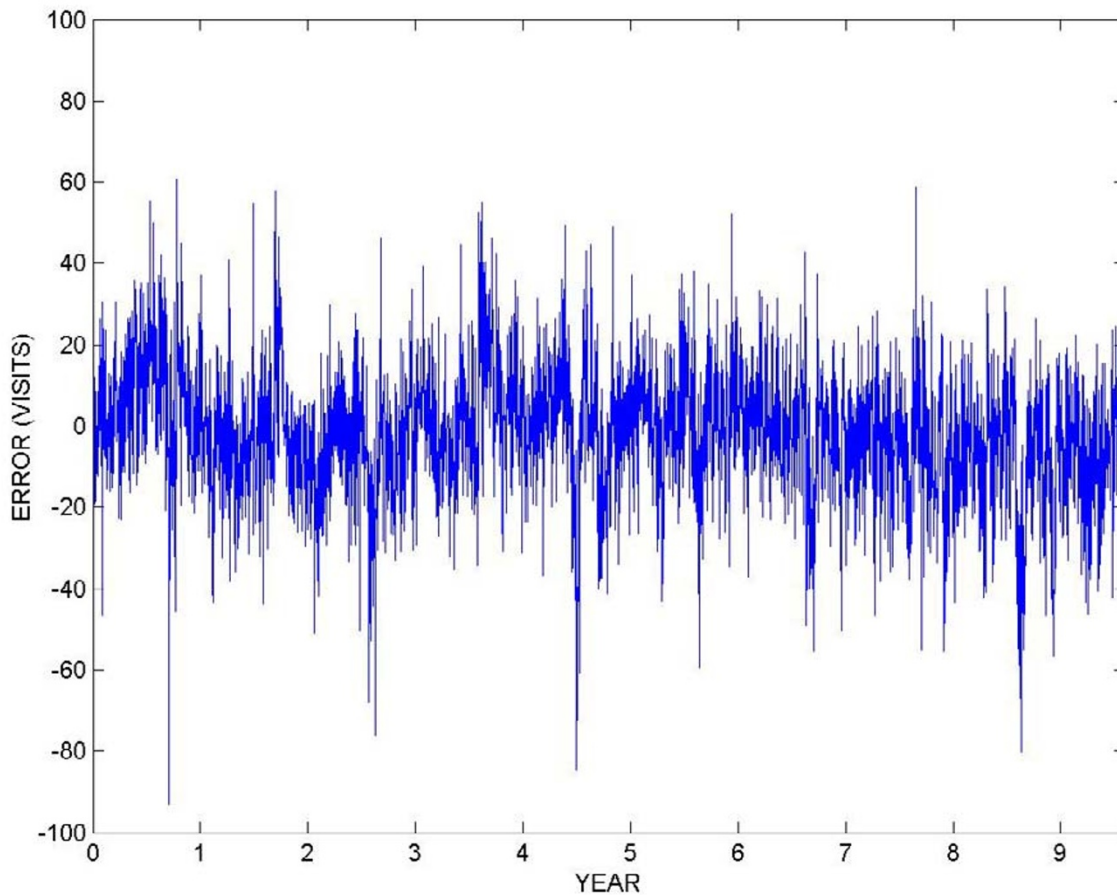
Figure 5
Forecast errors Forecast errors before (Figure 5) and after (Figure 6) ARIMA modeling of residuals.

Table 1: Quality of fit measured as mean absolute percentage error for various forecasting models on the training dataset.

	Simple Mean	Seasonal	Trimmed Seasonal	ARIMA
Total Visits %	13.17	9.16	8.96	8.80
Respiratory Visits %	35.29	22.69	21.74	18.35

Model development
The calculated MAPEs for various models of increasing sophistication for both overall and respiratory-related utilization volume in the training dataset are shown in Table 1.

An ARIMA(2,0,1) model for overall volume yielded a MAPE of 9.37%. This means that that on average, the forecasts of the model were to within an average of 13 visits of the actual daily volume, which averaged 137 visits per day.
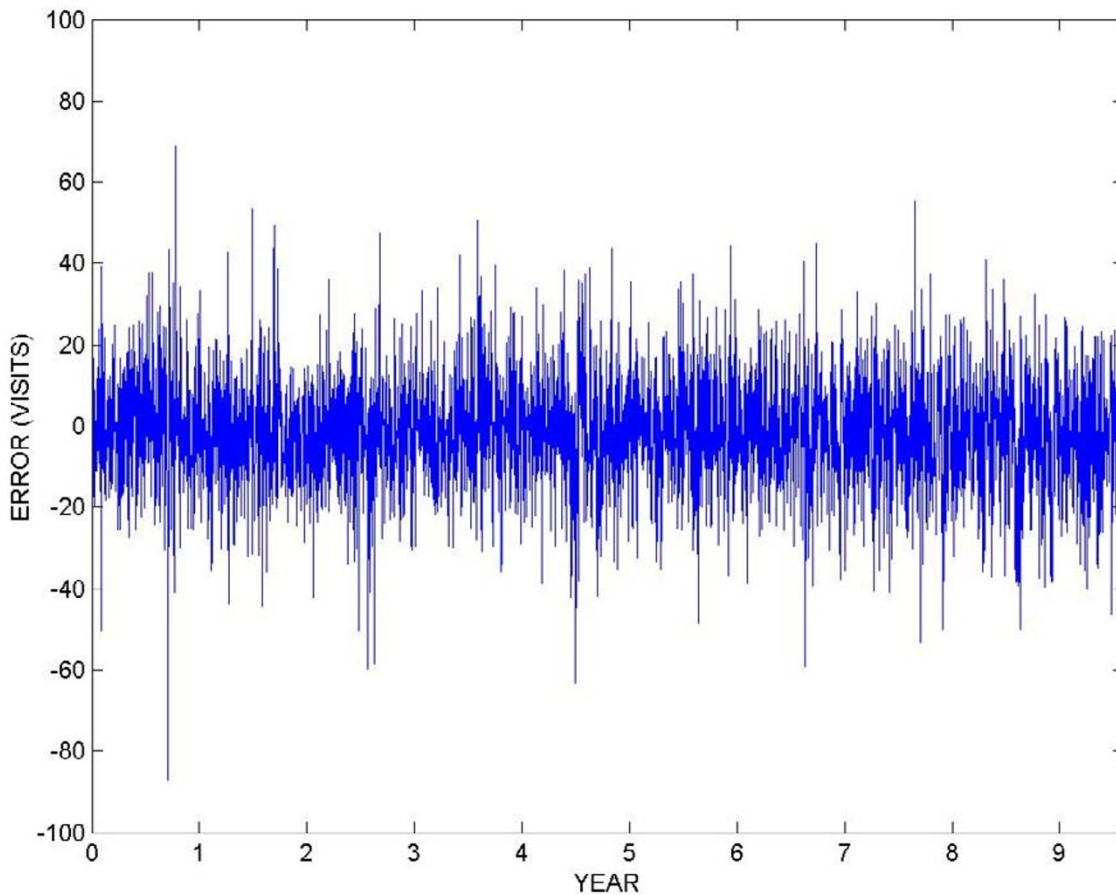
Figure 6
Forecast errors Forecast errors before (Figure 5) and after (Figure 6) ARIMA modeling of residuals.

For respiratory-related visits, An ARIMA(1,0,1) model yielded an overall MAPE of 27.54%. This means that that on average, the forecasts of the model were to within 13 visits of an average of 48 visits per day.

Figures 5 and 6 show the forecast error before and after ARIMA modeling of residuals. The error of the ARIMA model is close to white noise, meaning that the ARIMA model is successful in accounting for most of the significant autocorrelations present in the residuals.

Model validation
The MAPE values of the forecasts generated by the various models for the validation dataset are shown in Table 2. The ARIMA model performs best, followed by the

trimmed-mean seasonal model and the other more simple models.

The results of the validation experiments using simulated outbreaks are shown in Figure 7. In detecting the simulated outbreaks, the model was able to detect 100% of simulated 7-day, 30-visits-per-day outbreaks, while maintaining the benchmark specificity of 97%. Detection performance decreased with smaller simulated outbreaks, with sensitivity gradually decreasing to 99% for 25 visits-per-day, 94% for 20 visits-per-day, 78% for 15 visits-per-day, 57% for 10 visits-per-day, and 36% for 5 visits-per-day.
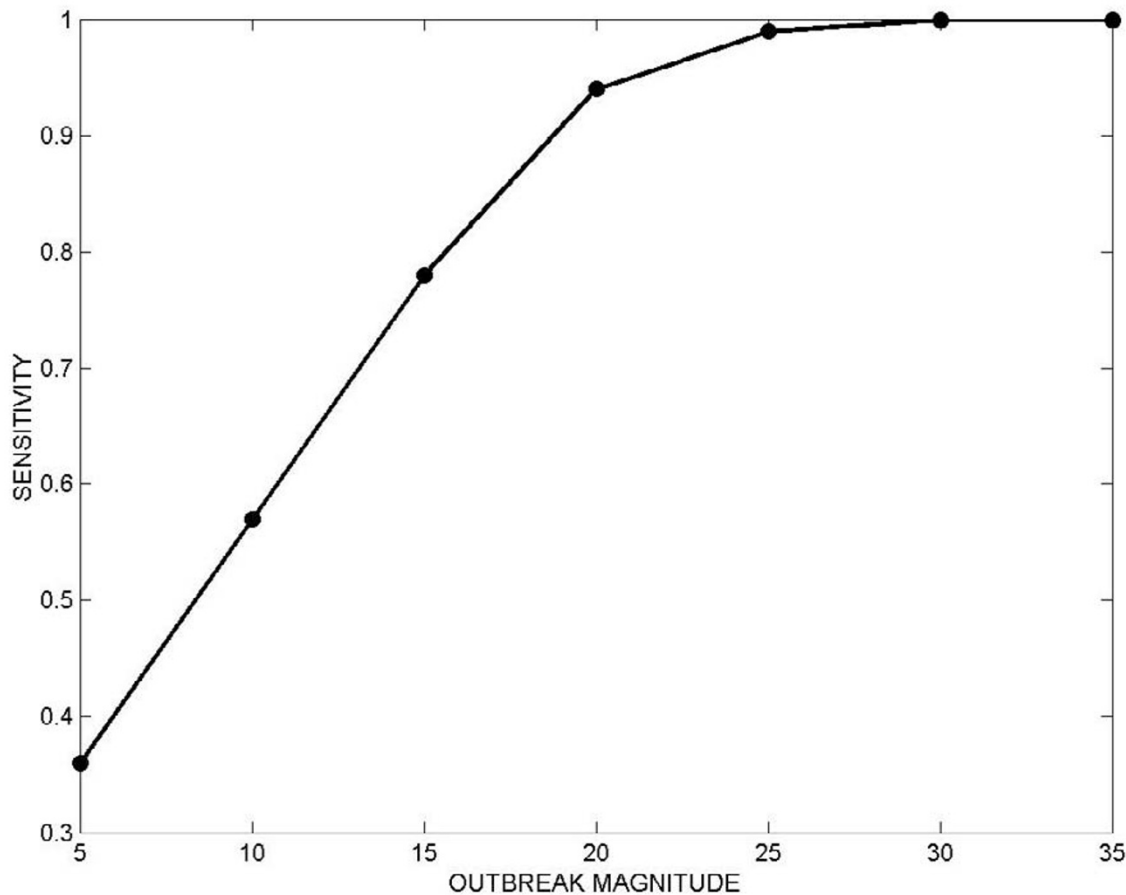
Figure 7
Detection Sensitivity The detection performance of the system when tested with simulated outbreaks of various magnitudes. Sensitivity measures how many of the 233 simulated outbreaks were detected by the system at each magnitude. The specificity was held at 97%, corresponding to roughly one false alarm per month.

Table 2: Forecast Accuracy measured as Mean Absolute Percentage Error for the various forecasting models on the validation dataset.

	Simple Mean	Seasonal	Trimmed Seasonal	ARIMA
Total Visits %	11.58	9.62	9.56	9.37
Respiratory Visits %	50.16	33.80	32.05	27.54

Discussion

The forecasting models developed here produced results with good accuracy, both in their fit to the training data, and more importantly, in their predictions of the validation data. The model of overall ED utilization also proved to be a solid basis for a detection system capable of detecting simulated outbreaks of different sizes with a high sensitivity and specificity. Sensitivity decreased with outbreak size.

Limitations of the model

The rate of respiratory-related visits was lower than the overall visit rate, and thus the model-based forecasts were less accurate for the respiratory-related visits. Forecasting accuracy might be improved by integrating multiple regional emergency departments into a single detection system, thus increasing the aggregate number of respiratory-related visits and allowing for more robust modelling of visit rates.

The accuracy of the models was also limited by the quality of the administrative chief complaint data used in the model. Using diagnostic ICD codes to identify the respiratory-related visits would probably improve the modelling accuracy. However, these diagnostic codes are not always available in real time for surveillance purposes.

Outbreaks of smaller magnitude were more difficult to detect since the few additional visits were likely to be masked by the inherent variability (noise) in the pattern of daily visits that is not fully captured in the model. For larger outbreaks, variability poses less of a problem since there is a larger signal-to-noise ratio.

Integrated alarms

Once forecasts have been generated, the appropriate alarm thresholds need to be set. The particular alarm strategy chosen affects the detection specificity and sensitivity of the detection system as a whole.

One of the primary benefits of ARIMA models is their ability to correct for local trends in the data – what has happened on the previous day is incorporated into the forecast of what will happen today. This works well, for example, during a particularly severe flu season, where prolonged periods of high visit rates are adjusted to by the ARIMA model, thus preventing the alarm from being triggered every day throughout the flu season.

However, if the ARIMA model "adjusts" to an actual outbreak instead of detecting it, a slowly spreading outbreak or attack might be missed because of this correction. This correction is most likely to affect detection of outbreaks occurring over several days, rather than those that occur suddenly. It is therefore also important to rely on the non-ARIMA model for outbreak detection.

To address this issue, we propose incorporating both a fixed seasonal model as well as a seasonal model with ARIMA residuals into a hybrid detection system. If an alarm is triggered by only the fixed seasonal model and not by the ARIMA model, the situation could be made clear by a textual explanation that accompanies the alarm. For example, a textual message might be displayed stating:

"Today's visit rates are significantly higher than expected for this day of the week and season. However, today's visit rates are in-line with the unseasonably high trend seen over the past few days."

Alternatively, when both models trigger an alarm, the textual explanation could read accordingly:

"Today's visit rates are significantly higher than expected for this day of the week and season. Furthermore, today's visit rates are also significantly higher than would be expected from the past few days."

By communicating the results of both models in this way, a more complete picture of healthcare utilization could be provided to the users of a detection system.

Conclusions

Based on extensive historical data, the models described here forecast both respiratory-related and overall pediatric ED volume with good accuracy, providing a solid basis for real-time surveillance and bioterrorism detection. A detection system based on the model of overall ED volume can detect simulated outbreaks with a high sensitivity and specificity.

From these models, an integrated and annotated alarms strategy that combines both historical and recent trends is proposed to provide a more complete picture of the current state of the public health. Using this integrated alarms strategy, a public health authority could better determine whether the overall and respiratory visit rates significantly exceed standardized thresholds. The systematic methodology described here can be generalized to other healthcare settings for developing biosurveillance systems that detect anomalous increases in healthcare utilization rates.

Competing interests

None declared.

Authors' contributions

BR conducted the study. BR and KM participated in its design and coordination and manuscript preparation. All authors read and approved the final manuscript.

Acknowledgements

References

1. From the Centers for Disease Control and Prevention **Considerations for distinguishing influenza-like illness from inhalational anthrax.** *JAMA* 2001, **286(20):**2537-9
2. Batal H **Predicting patient visits to an urgent care clinic using calendar variables.** *Acad Emerg Med* 2001, **8(1):**48-53

3. Bowdish GE **Using regression analysis to predict emergency patient volume at the Indianapolis 500 mile race.** *Ann Emerg Med* 1992, **21(10):**1200-3

4. Hagen JA **Associations between outdoor air pollutants and hospitalization for respiratory diseases.** *Epidemiology* 2000, **11(2):**136-40

5. Green MS and Kaufman Z **Surveillance for early detection and monitoring of infectious disease outbreaks associated with bioterrorism.** *Isr Med Assoc J* 2002, **4(7):**503-6

6. **AHRQ: IT/DSS can aid in bioterror response.** *Healthcare Benchmarks Qual Improv* 2002, **9(10):**46-8

7. Kohane IS **The contributions of biomedical informatics to the fight against bioterrorism.** *J Am Med Inform Assoc* 2002, **9(2):**116-9

8. Kun LG and Bray DA **Information infrastructure tools for bioterrorism preparedness. Building dual- or multiple-use infrastructures is the task at hand for state and local health departments.** *IEEE Eng Med Biol Mag* 2002, **21(5):**69-85

9. Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, McGinnis LF, Deerfield DW, Druzdzel MJ and Fridsma DB **The emerging science of very early detection of disease outbreaks.** *J Public Health Manag Pract* 2001, **7(6):**51-9General scientific

10. Morse SS **The vigilance defense.** *Sci Am* 2002, **287(4):**88-9

11. Goldenberg A, Shmueli G, Caruana RA and Fienberg SE **Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales.** *Proc Natl Acad Sci U S A* 2002, **99(8):**5237-40

12. Shannon M, Burstein J, Mandl K and Fleisher G **Usage of a web-based decision support tool for bioterrorism detection.** *Am J Emerg Med* 2002, **20(4):**384-5

13. Harcourt SE, Smith GE, Hollyoak V, Joseph CA, Chaloner R, Rehman Y, Warburton F, Ejidokun OO, Watson JM and Griffiths RK **Can calls to NHS Direct be used for syndromic surveillance?** *Commun Dis Public Health* 2001, **4(3):**178-82

14. Lazarus R, Kleinman K, Dashevsky I, Adams C, Kludt P, DeMaria A Jr and Platt R **Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events.** *Emerg Infect Dis* 2002, **8(8):**753-60

15. **Syndromic surveillance for bioterrorism following the attacks on the World Trade Center – New York City, 2001.** *MMWR Morb Mortal Wkly Rep* 2002, **51 Spec No:**13-5

16. Lewis M, Pavlin J, Mansfield J, O'Brien S, Boomsma L, Elbert Y and Kelley P **Disease outbreak detection system using syndromic data in the greater Washington DC area(1).** *Am J Prev Med* 2002, **23(3):**180

17. Lober WB **Roundtable on Bioterrorism Detection: Information System-based Surveillance.** *J Am Med Inform Assoc* 2002, **9(2):**105-15

18. Tsui FC, Wagner MM, Dato V and Chang CC **Value of ICD-9 coded chief complaints for detection of epidemics.** *Proceedings/ AMIA Annual Symposium.* 2001, 711-5

19. Espino JU and Wagner MM **Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness.** *Proceedings/AMIA Annual Symposium* 2001, 164-8

20. Barthell EN, Cordell WH, Moorhead JC, Handler J, Feied C, Smith MS, Cochrane DG, Felton CW and Collins MA **The Frontlines of Medicine Project: a proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism.** *Ann Emerg Med* 2002, **39(4):**422-9

21. **Matlab User's Guide.** *Mathworks Inc, Natick, MA*

22. Box GEP, Jenkins GM and Reinsel GC **Time Series Analysis: Forecasting & Control.** *Prentice Hall* 1994,

23. **SAS User's Guide** *SAS Institute Inc., Cary, NC*

24. Reis BY, Pagano M and Mandl KD **Using Temporal Context to Improve Biosurveillance.** *Proceedings National Academy of Sciences* **100(4):**1961-1965

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6947/3/2/prepub