Detection of outbreaks using laboratory data

An epidemiological perspective

David Buckeridge, MD PhD
McGill University, Montréal, Canada
Direction de santé publique de Montréal
Institut national de santé publique du Québec (INSPQ)

October 23, 2006
Outline

The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
Outline

The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
The Process of Automated Surveillance

Surveillance methodology and approaches to surveillance

The process of surveillance

All approaches to surveillance share some common principles. While some of the underlying methods used in public health surveillance have evolved considerably in recent years, the general approach to surveillance has remained relatively constant. At a fundamental level, surveillance aims to (1) identify individual cases, (2) detect population patterns in identified cases, and then (3) convey information to decision-makers about population health patterns (Fig. 3).

Identification of individual cases

The definition of a case for a surveillance system (Fig. 3, Step 1) has important implications for the design and performance of the system. In settings where a surveillance system is intended to follow cases of a well-understood disease, it may be possible to make the case definition highly specific. For example, public health agencies in many developed countries conduct routine surveillance for communicable diseases such as measles. Definitions of cases in these systems tend to rely upon highly specific diagnostic tests. As a result, communicable disease surveillance...

(Buckeridge & Cadieux, 2007)
Laboratory Surveillance Contexts

- **Within hospitals**
  - Nosocomial surveillance.
  - Adverse event detection (e.g., post-op infection).

- **Regional, National, International**
  - Notifiable disease surveillance.
  - Syndromic surveillance.
  - ...
Data Elements Used in Laboratory Surveillance

- **Incidence of test orders** (Ma et al, 2005)
  - Timely, sensitive signal.
  - Reflects clinical suspicion.
- **Incidence of positive results**
  - Results analyzed by organism or host characteristics.
  - **Organism – type, subtype** (e.g., Bender 2001 for Salmonella), antimicrobial resistance.
  - **Person – Demographics, Location, Comorbidities.**
Purposes of Laboratory Surveillance

- Monitoring antimicrobial resistance.
- Monitoring disease management
  - Diabetes – Haemoglobin A1C.
  - Cardiovascular disease – Lipids.
- ... 
- Detecting disease cases.
- Detecting disease outbreaks.
Outline

The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
Determinants of Disease Outbreak Detection

Outbreak Factors
- Endemic Disease
  - incidence
  - variation / shape
- Epidemic Disease
  - incidence
  - variation / shape
  - timing of onset

Data Collection
- data source(s)
- sampling frame, method, frequency
Case Detection
- case definition
- case classification algorithm
- processing frequency
Outbreak Detection
- algorithm
- threshold
- analysis frequency
Public Health Action
- interpretation of analysis
- investigation
- response

System Factors

(Buckeridge, 2007)
Outline

The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
There is Strong Evidence that Automated Case Reporting is Effective

- Many studies support automated case-reporting within a reportable disease context. (e.g., Effler, 1999)
- Result reporting is faster, more representative and more complete.
- Reporting delay decreases from approximately 5 days to approximately 1 day.
Simple Case Definitions are Used in Practice, but More Information is Better

- Single code matching on diagnosis and/or result is the usual approach.
  - Reports tests, not truly cases.
  - Inconsistent coding is a practical problem. (Overhage, 2001)
- In a hospital setting, automated lab surveillance has reasonable sensitivity compared to review of patients – Patient review uses additional sources.
- When lab data are combined with other indicators,† sensitivity increases, but impact on specificity (false positives) is not well-defined.

†Prescriptions, risk factors such as age, time in hospital, alternate diagnoses
Outline

The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
The Appropriate Analytic Approach Varies

- The statistical algorithms should be selected to match the data and the likely outbreak signal. (Buckeridge, 2005)
- Resolution of data
  - Temporal frequency.
  - Spatial precision.
  - Other covariates.
- Type of outbreak signal
  - Spatially clustering.
  - Slowly or rapidly increasing.
  - Focussed in high-risk populations.
Temporal Control Charts are Used Commonly

\[ y_t = \max \left( 0, y_{t-1} + \left( \frac{x_t - \mu_t}{\sigma_t} - k \right) \right) \]

- Cumulative sum (cusum) method works well for data grouped by weeks or months.
- Statistical Process Control (SPC) Chart methods are taken from manufacturing
  - System is ‘in control’ or ‘out of control’.
  - Other methods include Shewhart, Exponentially Weighted Moving Average (EWMA).
- Also, ad hoc smoothing methods (e.g., Stern 1999).
Using Laboratory-Based Surveillance Data for Prevention: An Algorithm for Detecting *Salmonella* Outbreaks

By applying cumulative sums (CUSUM), a quality control method commonly used in manufacturing, we constructed a process for detecting unusual clusters among reported laboratory isolates of disease-causing organisms. We developed a computer algorithm based on minimal adjustments to the CUSUM method, which cumulates sums of the differences between frequencies of isolates and their expected means; we used the algorithm to identify outbreaks of *Salmonella* Enteritidis isolates reported in 1993. By comparing these detected outbreaks with known reported outbreaks, we estimated the sensitivity, specificity, and false-positive rate of the method. Sensitivity by state in which the outbreak was reported was 0%(0/1) to 100%. Specificity was 64% to 100%, and the false-positive rate was 0 to 1.

(Hutwagner, 1997)
Time-Series Methods in Surveillance

- Time-series forecast methods are more technically demanding, but useful for data with regular temporal cycles, such as day of the week.
- A regression model that accounts for temporal autocorrelation is used to forecast expected values.
- The observed value is compared to the forecast and the residual or difference is used to detect outbreaks.
- Time-series methods also provide a framework for combing data from laboratories and other sources in a hybrid surveillance scheme.
A MONITORING SYSTEM FOR DETECTING ABERRATIONS IN PUBLIC HEALTH SURVEILLANCE REPORTS

G. DAVID WILLIAMSON\textsuperscript{1,*} AND GINNER WEATHERBY HUDSON\textsuperscript{2}

\textsuperscript{1}Epidemiology Program Office, MS K73, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A.
\textsuperscript{2}Drexel University, College of Business and Administration, Matheson Hall, 32nd and Market Streets, Philadelphia, PA 19104, U.S.A.

SUMMARY

Routine analysis of public health surveillance data to detect departures from historical patterns of disease frequency is required to enable timely public health responses to decrease unnecessary morbidity and mortality. We describe a monitoring system incorporating statistical ‘flags’ identifying unusually large increases (or decreases) in disease reports compared to the number of cases expected. The two-stage monitoring system consists of univariate Box–Jenkins models and subsequent tracking signals from several statistical process control charts. The analyses are illustrated on 1980–1995 national notifiable disease data reported weekly to the Centers for Disease Control and Prevention (CDC) by state health departments and published in CDC’s Morbidity and Mortality Weekly Report. Published in 1999 by John Wiley & Sons, Ltd. This article is a U.S. Government work and is in the public domain in the United States.
Forecasting Expected Cases of Hepatitis A

Figure 5. Predicted and actual number of reported cases of hepatitis A, by week, United States, 1993–1995 (the 1993–1995 predicted numbers are model estimates, the 1994–1995 predicted numbers are rolling forecasts and the actual number of cases are provisional data from 9 January 1993 to 13 May 1995).

Figure 6. Shewhart control chart for the number of reported cases of hepatitis A, by week, United States, 1989. No significantly high values were identified for the 1989–1990 forecasted national hepatitis A disease report series.

For 1989 hepatitis A forecasted data, monitoring results varied by state with no evident relationships for statistically high values among the three states or with the national disease monitoring results. High aberrations were detected by at least one control chart for each of the states.

(Williamson, 1999)
The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
  Case detection
  Outbreak detection

Evaluation of Outbreak Detection
The Metrics of Outbreak Detection

- Sensitivity, Specificity, Receiver Operating Characteristics (ROC) curves, and Area Under the Curve (AUC) are used to measure accuracy.

- Timeliness, Activity Monitoring Operating Characteristics (AMOC) curves, TROC (Timeliness Receiver Operating Characteristics) surface, and Volume Under the Surface (VUS) are used to assess the delay until detection.

- Other measures, such as the proportion of infections averted, are used less frequently.
Receiver Operating Characteristics (ROC) Curves

False alarms per month (1 analysis per day)

- 100 infected
- 500 infected
- 1000 infected
- 5000 infected

Specificity vs. Sensitivity

1 - Specificity

0.0 0.2 0.4 0.6 0.8 1.0

100 infected
500 infected
1000 infected
5000 infected
Measuring Accuracy and Timeliness of Outbreak Detection

Timeliness–Receiver Operating Characteristic Surface
Approaches to Evaluating Outbreak Detection

- Real outbreaks, if available, are preferable.
- The superimposition of simulated outbreaks onto real data is a good alternative.
A Superimposed Outbreak

Respiratory Visits

Real data
Real + Simulated data
Forecast
Upper 95% CI
Cusum alarm

Respiratory Visits

Mon Tue Wed Thu Fri Sat Sun Mon Tue Wed Thu Fri Sat Sun Mon Tue Wed Thu Fri Sat Sun
Summary

- Laboratory surveillance can be performed in various contexts for a range of reasons.
- Automated case reporting improves completeness and timeliness.
- A range of outbreak detection methods exists for different types of data.
- Metrics exist for accuracy and timeliness of detection.
References