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

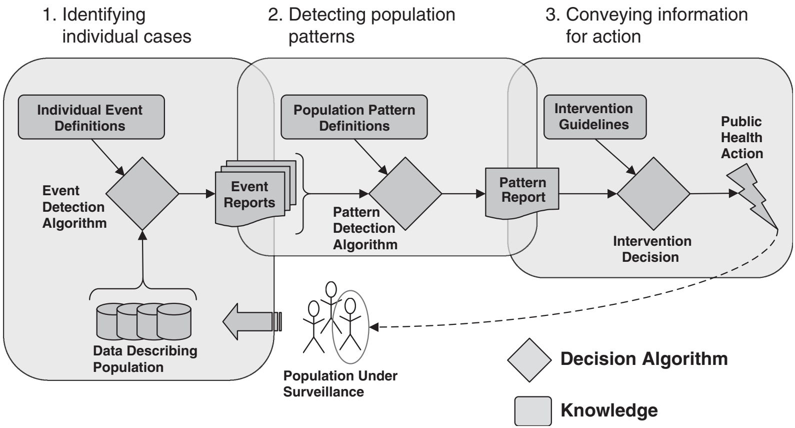
The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance
- Case detection
- Outbreak detection

Evaluation of Outbreak Detection

Outline

The Process of Automated 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

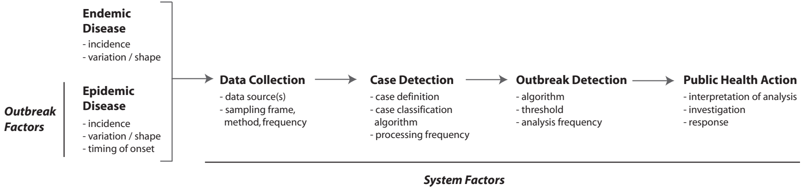
The Context of Laboratory Surveillance

Evidence and Methods for Automated Surveillance

- Case detection
- Outbreak detection

Evaluation of Outbreak Detection

Determinants of Disease Outbreak Detection



(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 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.

Using Time-Series Methods for Reportable Disease Surveillance

STATISTICS IN MEDICINE
Statist. Med. **18**, 3283–3298 (1999)

A MONITORING SYSTEM FOR DETECTING ABERRATIONS IN PUBLIC HEALTH SURVEILLANCE REPORTS†

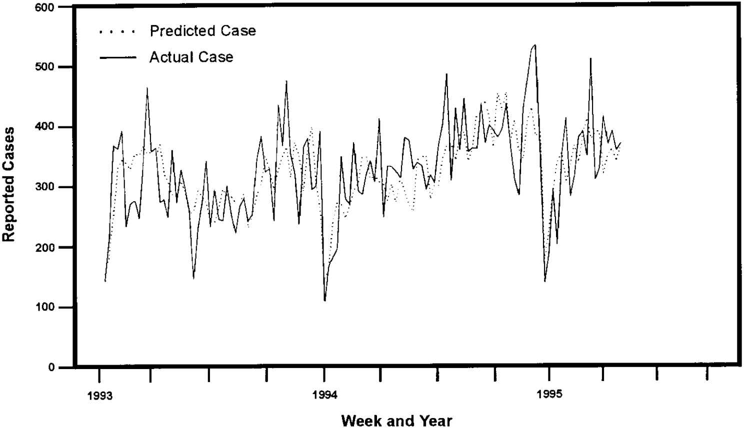
G. DAVID WILLIAMSON[1]* AND GINNER WEATHERBY HUDSON[2]

[1] *Epidemiology Program Office, MS K73, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A.*
[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



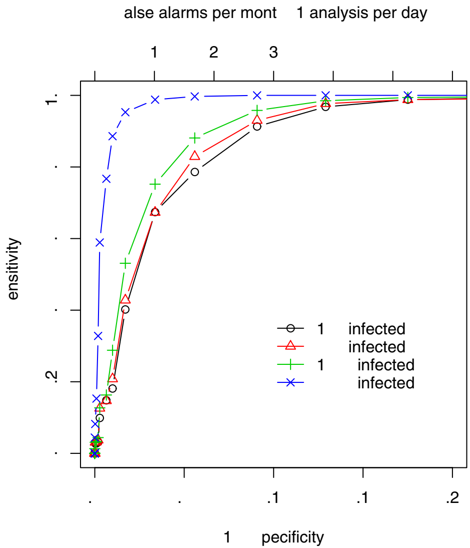
(Williamson, 1999)

Outline

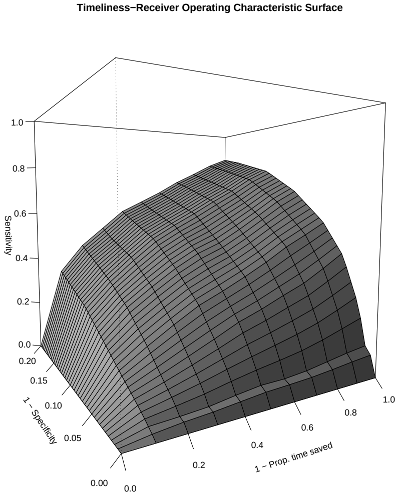
The Metrics of Outbreak Detection

- 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



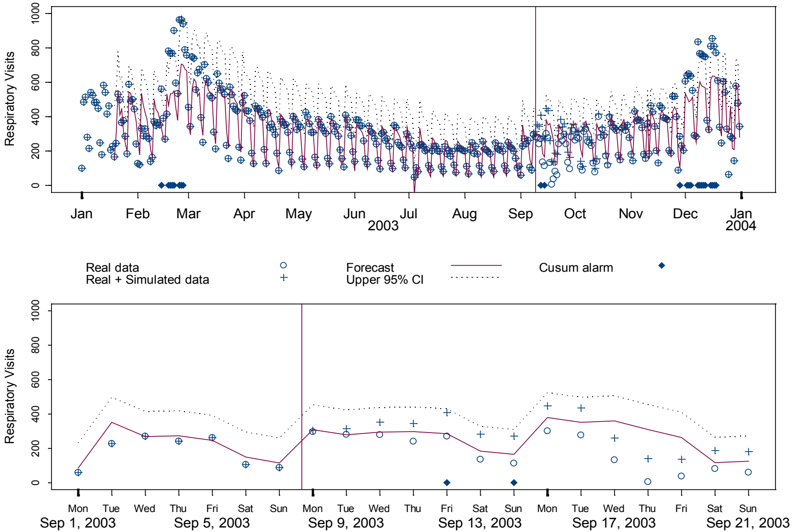
Measuring Accuracy and Timeliness of Outbreak Detection



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



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

→ Bender JB et. al. Use of molecular subtyping in surveillance for Salmonella enterica serotype typhimurium. N Engl J Med. 2001 Jan 18;344(3):189-95.

→ Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. J Biomed Inform. 2005 Apr;38(2):99-113.

→ Buckeridge DL. Outbreak Detection through Automated Surveillance: A Review of the Determinants of Detection. J Biomed Inform. 2007. To Appear.

→ Buckeridge DL and Cadieux G (2007). Surveillance for Newly Emerging Viruses. In: Tabor E (Ed). Perspectives in Medical Virology: Emerging Viruses in Human Populations. London: Elsevier.

→ Effler P, Ching-Lee M, Bogard A, Ieong MC, Nekomoto T, Jernigan D. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. JAMA. 1999 Nov 17;282(19):1845-50.

→ Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis. 1997 Jul-Sep;3(3):395-400.

→ Ma H, Rolka H, Mandl K, Buckeridge DL, Endy T, Fleischauer A, Pavlin J. Lab order data implementation in BioSense early outbreak detection system. In: Syndromic Surveillance: Reports from a National Conference, 2004. MMWR 2005;54(Suppl):27-30.

→ Overhage JM, Suico J, McDonald CJ. Electronic laboratory reporting: barriers, solutions and findings. J Public Health Manag Pract. 2001 Nov;7(6):60-6.

→ Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis. Epidemiol Infect. 1999 Feb;122(1):103-10.