Stoto, Schonlau & Mariano (2004). Syndromic Surveillance: Is it Worth the Effort? Chance, 17 (1), 19-24.

Bahjat F. Qaqish
(Bahjat_Qaqish@unc.edu)

The University of North Carolina at Chapel Hill

Conclusuions

- ... the benefits of SS have not yet been established
- SS offers the potential for early warning of bioterrorism attacks and natural disease outbreaks, but this potential is limited and the cost can be high
- The potential beyond bioterrorism *may* make syndromic surveillance worth the effort

Definitions

Syndromic Surveillance (SS) involves the statistical analysis of data from emergency rooms, over the counter medication sales and absenteeism records for early symptoms and signs of bioterrorism agents.

Sensitivity = pr(alarm | outbreak)

Specificity = pr(no alarm | no outbreak)

False Positive Rate = pr(alarm | no outbreak)
= 1− Specificity

False positive rate ≠ pr(alarm but no outbreak)

Issues

- Tradeoffs:
  1. Sensitivity versus specificity
  2. Timeliness
- Four questions:
  1. What kinds of attacks are SS systems reasonably able to detect?
  2. Can more sophisticated statistical methods do better than simple methods?
  3. Should SS be integrated into public health practice?
  4. How can SS systems help meet other health needs?
- First 2 questions - simulation
- Last 2 questions - discussion

Simulation Study

Data: January 1998 - June 2001, $N \approx 1000$ days

George Washington University Hospital daily number of cases of influenza-like illness

Average 3/day, peak 20 during flu season, Figure 1

Data were seeded with outbreaks, then subjected to different detection algorithms

Two types of outbreaks, Figure 2:

- Fast outbreak over 3 days (3,6,9)
  3 cases on day 1, 6 on day 2, 9 on day 3
- Slow outbreak over 9 days (1,1,1,2,2,2,3,3,3,)
  1 case on day 1, 1 on day 2, $\cdots$, 3 on day 9

Original data: $x_1, \cdots, x_{N+8}$

Generate N sequences y with outbreaks

To generate the ith fast outbreak: Set $y_j = x_j$ for $1 \leq j \leq N$ except

$$y_i = x_i + 3$$

$$y_{i+1} = x_{i+1} + 6$$
$$y_{i+2} = x_{i+2} + 9$$

Separate analysis of winter flu season vs rest of the year

Four detection algorithms

1. Shewhart(h): Alarm on day t if $y_t > h$
2. EWMA (h, λ):

   $$z_t = \lambda y_t + (1 - \lambda) z_{t-1}$$

   Alarm on day t if $z_t > h$
3. CUSUM(h, k):

   $$c_t = \max(0, y_t - \mu - k + c_{t-1})$$

   Alarm on day t if $c_t > h$
   $\mu :=$ daily average ?=3
4. Mean-adjusted CUSUM (h, λ, k):

   $$c_t = \max(0, y_t - z_{t-1} - k + c_{t-1})$$

   Alarm on day t if $c_t > h$

h chosen to give 1% false positive rate over the 3 years (10 alarms on the original data of 1000 days?)

k, λ chosen to maximize true positive rate with 1% 1% false positive rate over the 3 years

Results

Fast outbreak over three days (3,6,9),
Figure 3A

- Outside flu season:
  - 5-10% on day 1
  - 45-55% on day 2
  - 100% on day 3
- In flu season:
  - All $\leq$ 20% on day 3

Results

Slow outbreak over 9 days (1,1,1,2,2,2,3,3,3,), Figure 3B

- Outside flu season:
  - Shewhart 5% on day 9
  - EWMA, CUSUMs 50% on day 9
- In flu season:
  - All $\leq$ 10% on day 9

Results Summary

Outside flu season:
Fast outbreak: 2 days for 50%, all 4 methods Slow outbreak: 9 days for 50%, EWMA and CUSUMs

In flu season:
All poor

Areas for Improvement?

- Assumptions about background (normal) level and patterns of symptoms:
  Filtering, seasonality, day/week/month effects.
  But, flu season varies in size and timing.
- Targeting specific departures from normal:
  e.g. Looking for fever and rash only can be more sensitive to smallpox, but not to agents with different symptoms.
  Geographic analysis.
- Algorithm (statistical method):
  Data pooling over multiple ER's and hospitals can lose localized increases.
  Can analyze separately and pooled, but mutliple testing increases false positive rate.

Integrating SS and Public Health

- The value of SS depends on how well it is integrated into public health systems
- Alarms raised by SS must be investigated. Costly.
- Causes of alarm may be: pharmacies havng a sale, skipping school on a nice spring day
- Identification data: Need to know who got the flu-like illnes, not just the number
- HIPAA and legal issues
  Sharing health data on the scale contemplated by many SS systems is simply not allowed under the regulations
  Exemption for *reportable* diseases.
- Available control startegies expensive, controversial, efficacy unkown.
- Making decisions could take days or weeks.
- Agents that spread versus those that don't

Is It Worth the Effort?

- Evaluation tools needed (e.g. Using natural disease outbreaks such as West Nile Virus and SARS).
- Multiple data streams (multiple time series).
- Spatial
- ... the benefits of SS have not yet been established
- SS offers the potential for early warning of bioterrorism attacks and natural disease outbreaks, but this potential is limited and the cost can be high
- The potential beyond bioterrorism *may* make syndromic surveillance worth the effort