CUSUM Techniques to Identify "Flu" Outbreaks

Michael A. Stoto, PhD
RAND
Janice Matheson, MPH
Virginia B. Foster, MPH, PhD
George Washington University

The problem

Knowing when the influenza season has started can allow health departments to

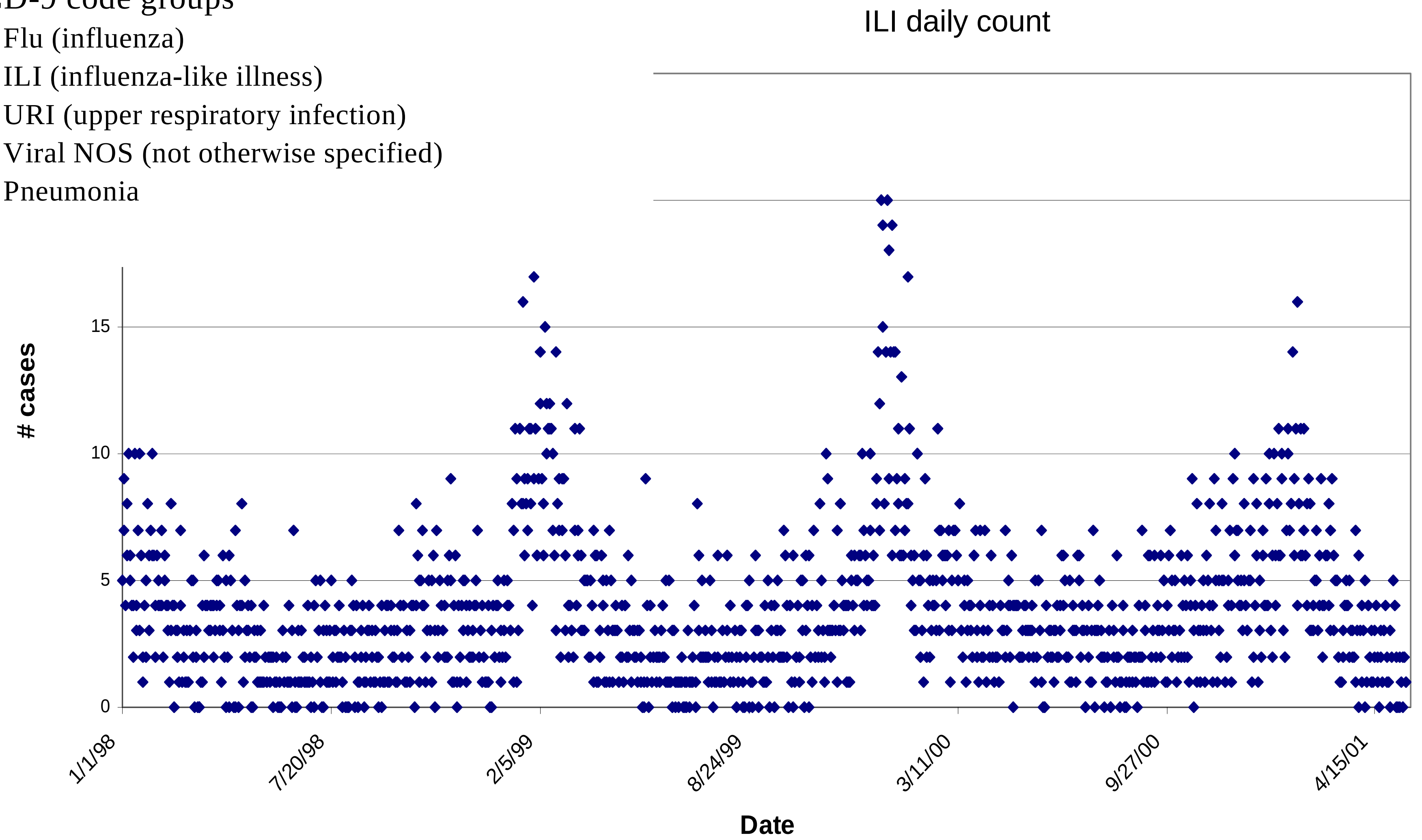
- Alert physicians and the public
- Allocate resources
- Start laboratory work to determine antigenic type

Hard to determine when the season has started

- At the community or facility level, reports of influenza and flu-like symptoms are infrequent, even at the start of the flu season

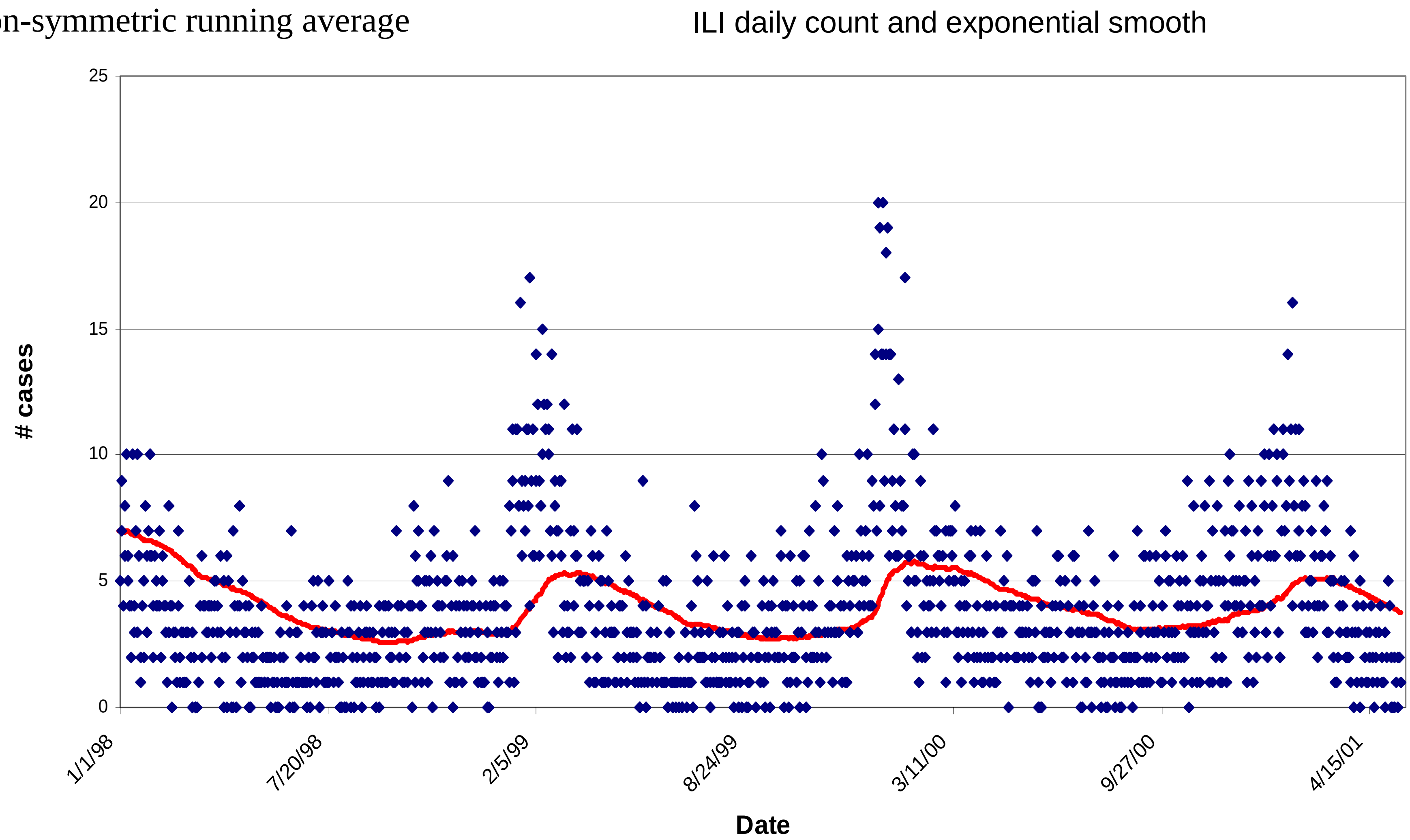
Data

- George Washington University Hospital Emergency Department
  - From Jan. 1, 1998 trough May 15, 2001
  - 3.5 flu seasons: Nov. 1 through May 15
- ICD-9 code groups
  - Flu (influenza)
  - ILI (influenza-like illness)
  - URI (upper respiratory infection)
  - Viral NOS (not otherwise specified)
  - Pneumonia



Exponential smoothing

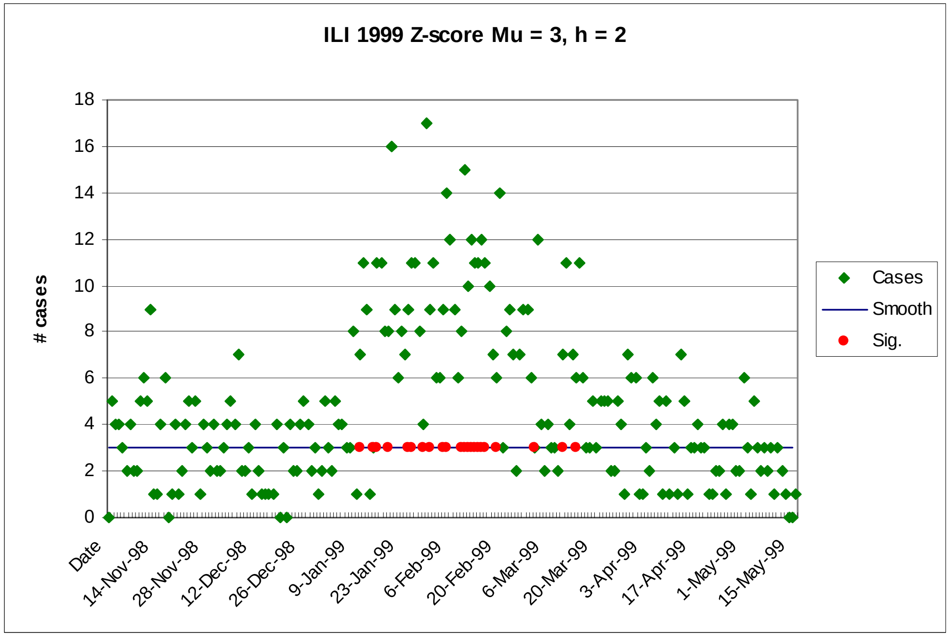
- Set w between
  - 0 (more smoothing) and
  - 1 (less smoothing)
- $y_t = wx_t + (1-w)y_{t-1}$
- non-symmetric running average

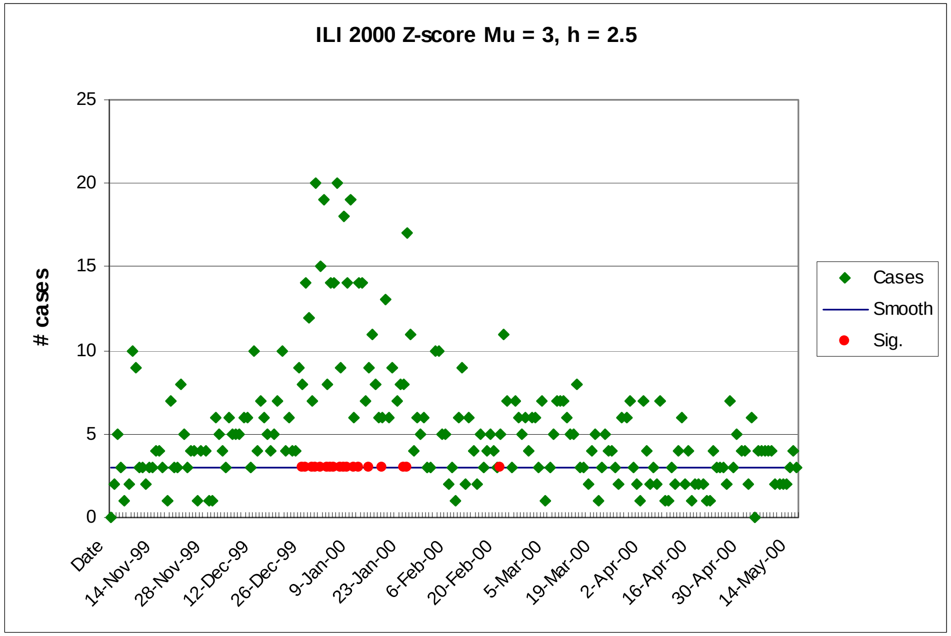


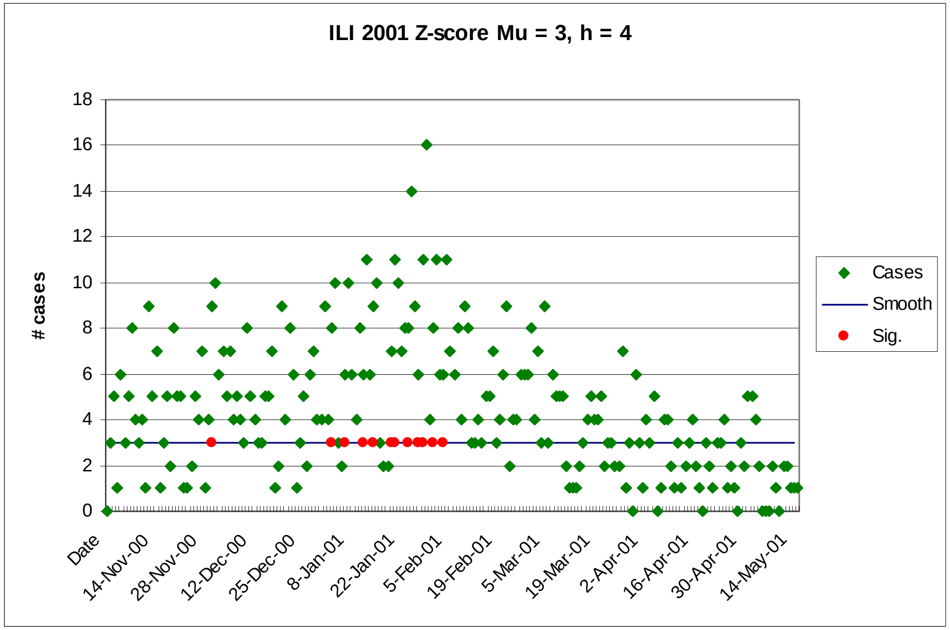
Standard outbreak detection method – Z-scores

- Estimate μ, $\sigma = \sqrt{\mu}$ (Poisson distribution)
  - $Z_t = (x_t - \mu)/\sigma$
  - Set h
  - "Outbreak" if $Z_t > h$
- Drawbacks
  - Many false positives and negatives
  - Optimal parameters vary by year
    - μ: 2 to 3
    - h: 2 to 4

Red dot indicates significant elevation

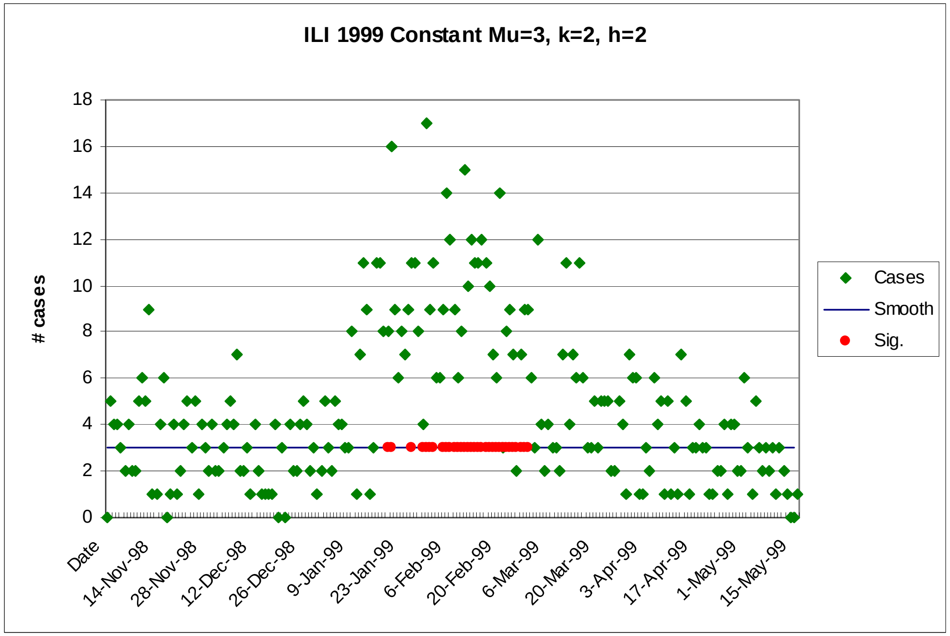


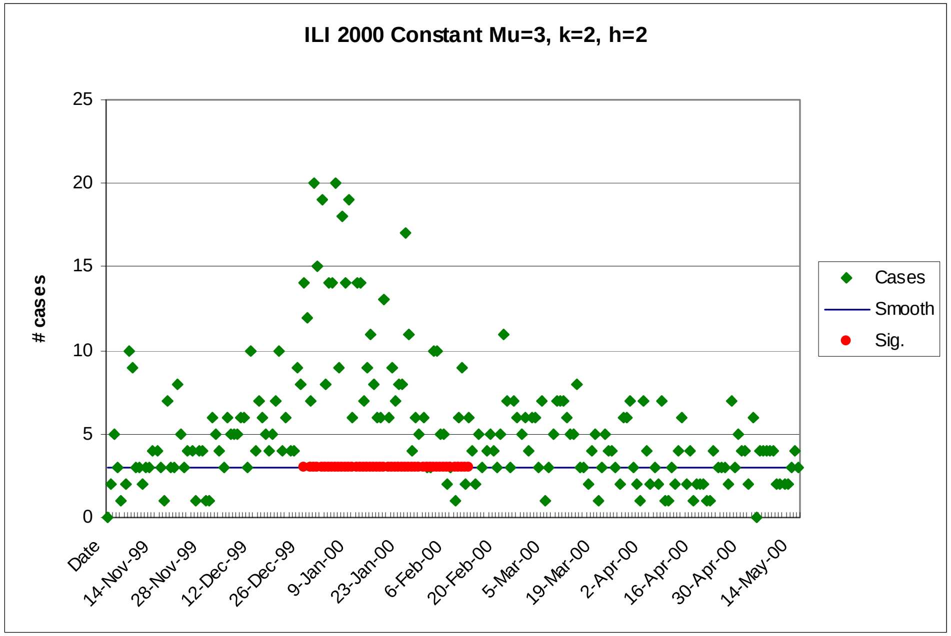


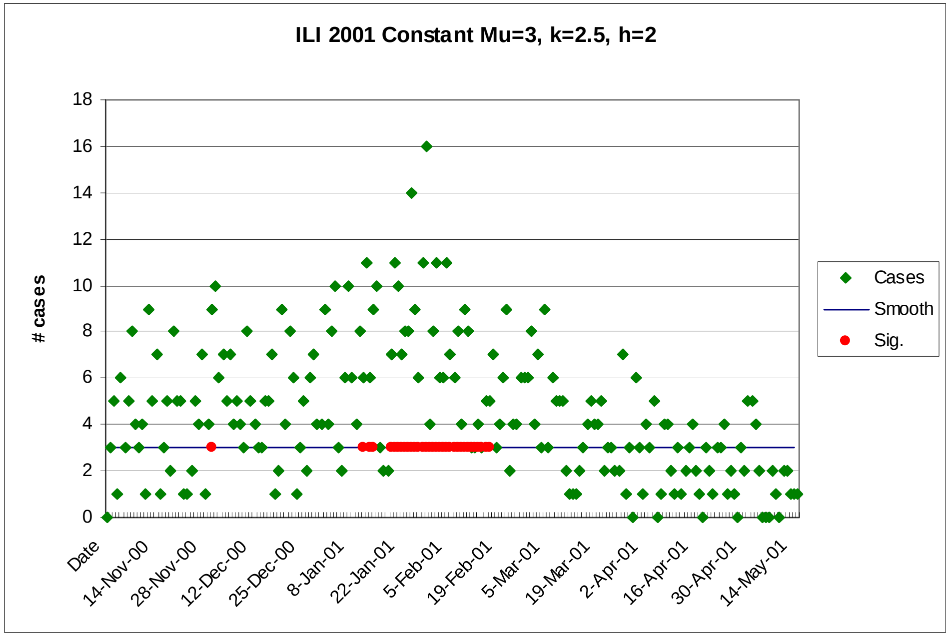


Alternative 1: CUSUM with constant background level

- Method
  - Estimate μ, σ = √μ (Poisson distribution)
  - $Z_t = (x_t - \mu)/\sigma$
  - Set k, h
  - $S_t = \text{Max}[S_{t-1} + Z_t - k, 0]$
  - "Outbreak" if $S_t > h$
- Comparison
  - Fewer false positives and negatives
  - Optimal parameters vary by year
    - k: 0.5 to 2.5



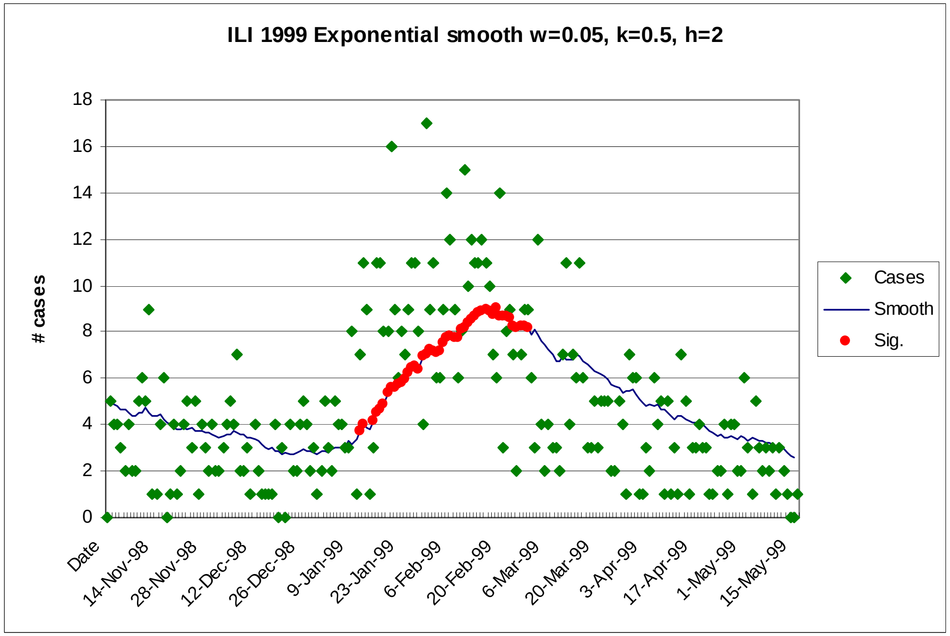


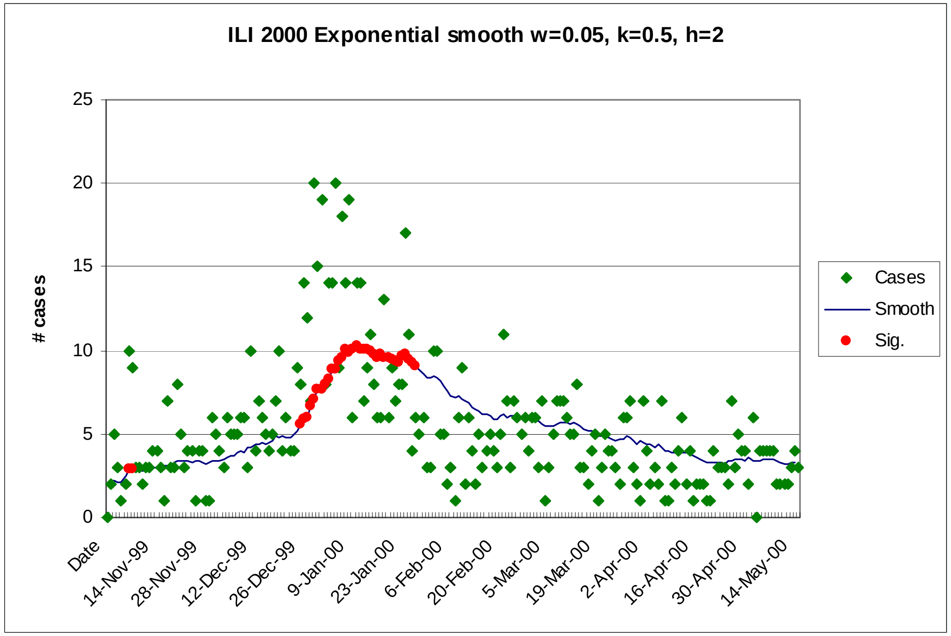


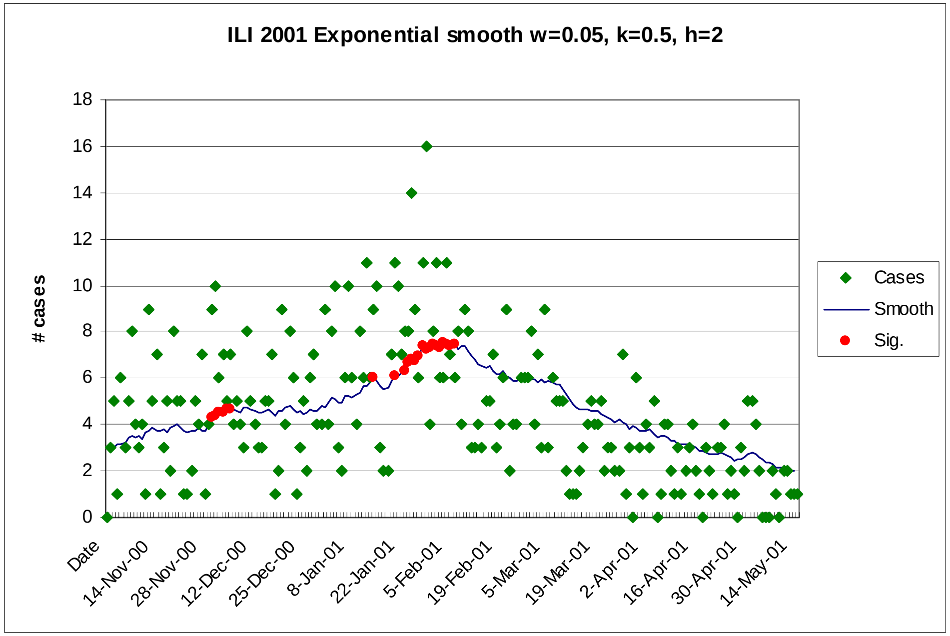
Red dot indicates significant CUSUM

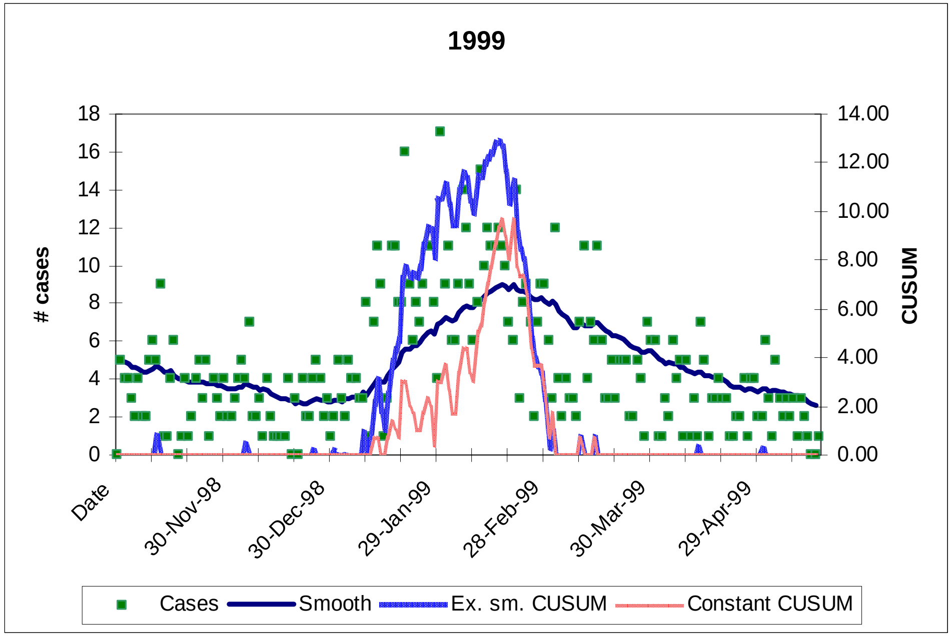
Alternative 2: CUSUM with exponential smoothing background level

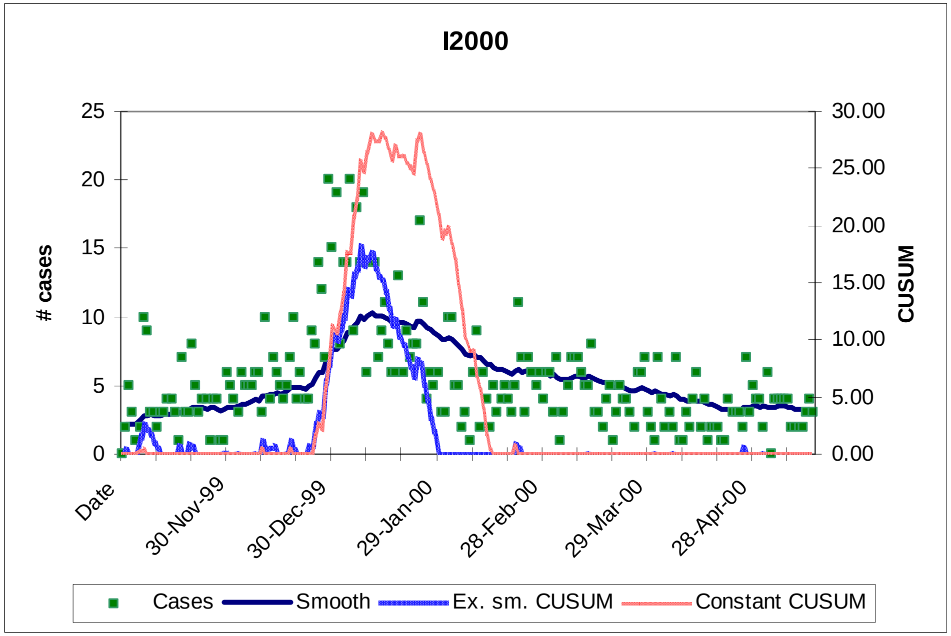
- Method
  - Set w, k, h, estimate σ from data
  - Calculate $y_t = wx_t + (1\text{-}w)y_{t\text{-}1}$
  - $Z_t = (x_t - y_t)/\sigma$
  - $S_t = \text{Max}[S_{t-1} + Z_t - k, 0]$
  - "Outbreak" if $S_t > h$
- Comparison
  - Earlier detection of '99 epidemic
  - Optimal parameters do not vary by year

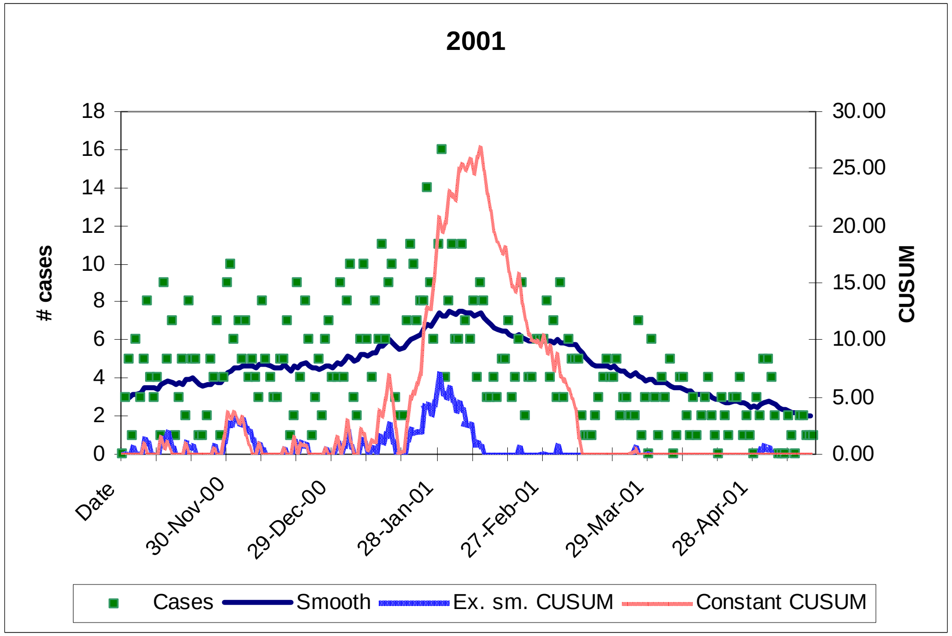








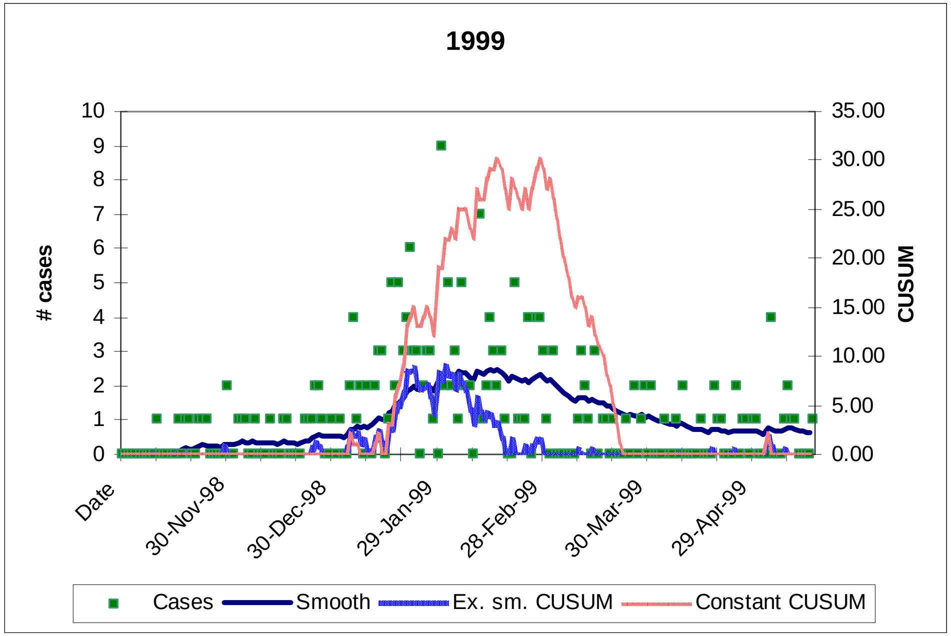


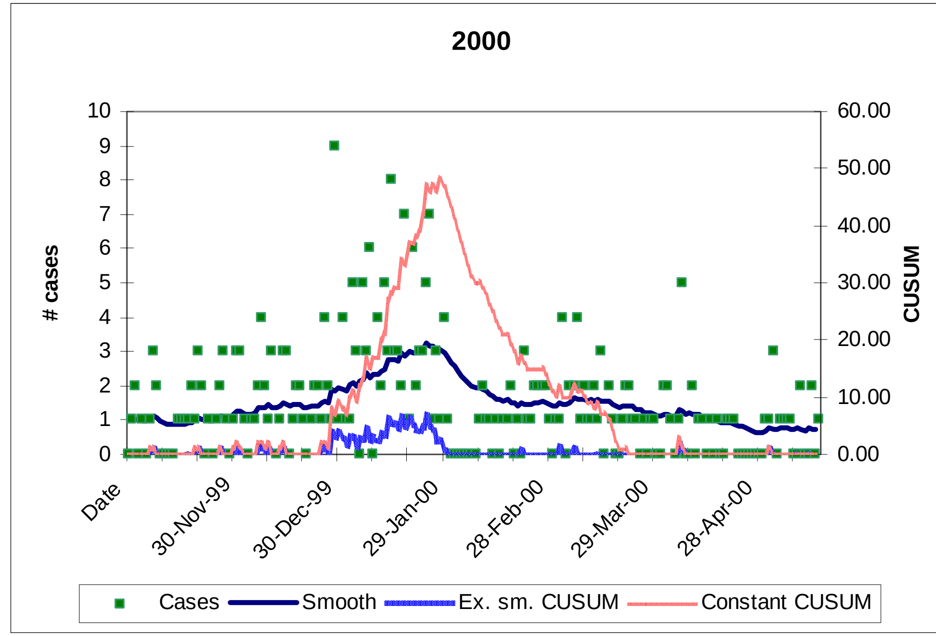


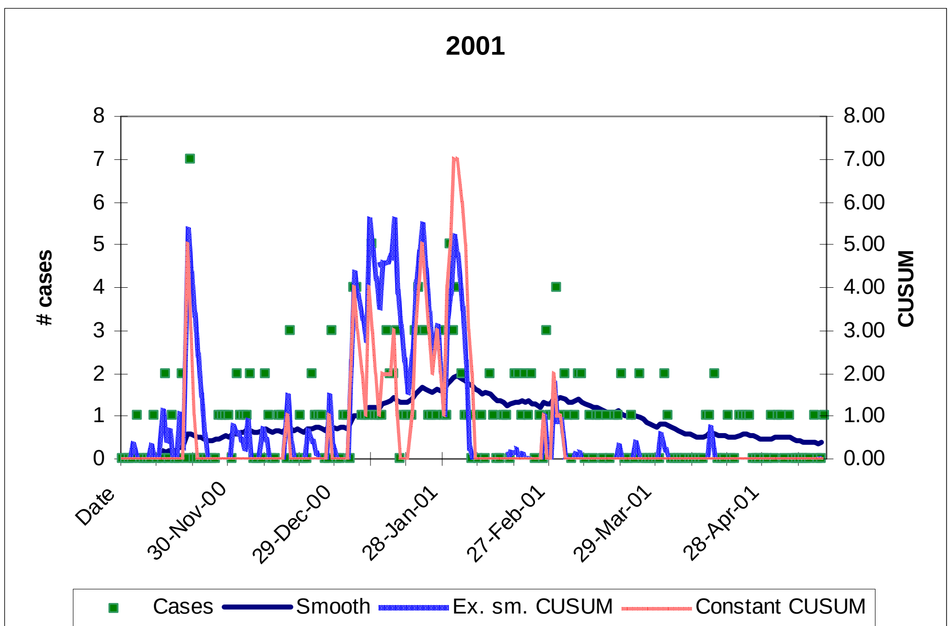
Constant vs. exponential smoothing CUSUM for ILI

- Set k=0.5 for ExSm and k= 2 for constant, and h=2, all years
- First date of ILI outbreak

Year	Start of "outbreak"				
	CUSUM		Regional/ widespread	Sentinal MDs	P&I Deaths
	Constant	Exp. Sm.			
1999	Jan 20	Jan 12	Jan 16	Jan 23	Jan 30
2000	Dec 27/29	Dec 26	Dec 18	Dec 25	Nov 27
2001	Dec 1/Jan 13	Dec 1/Jan 25	Nov 18/Jan 6	Jan 20	X





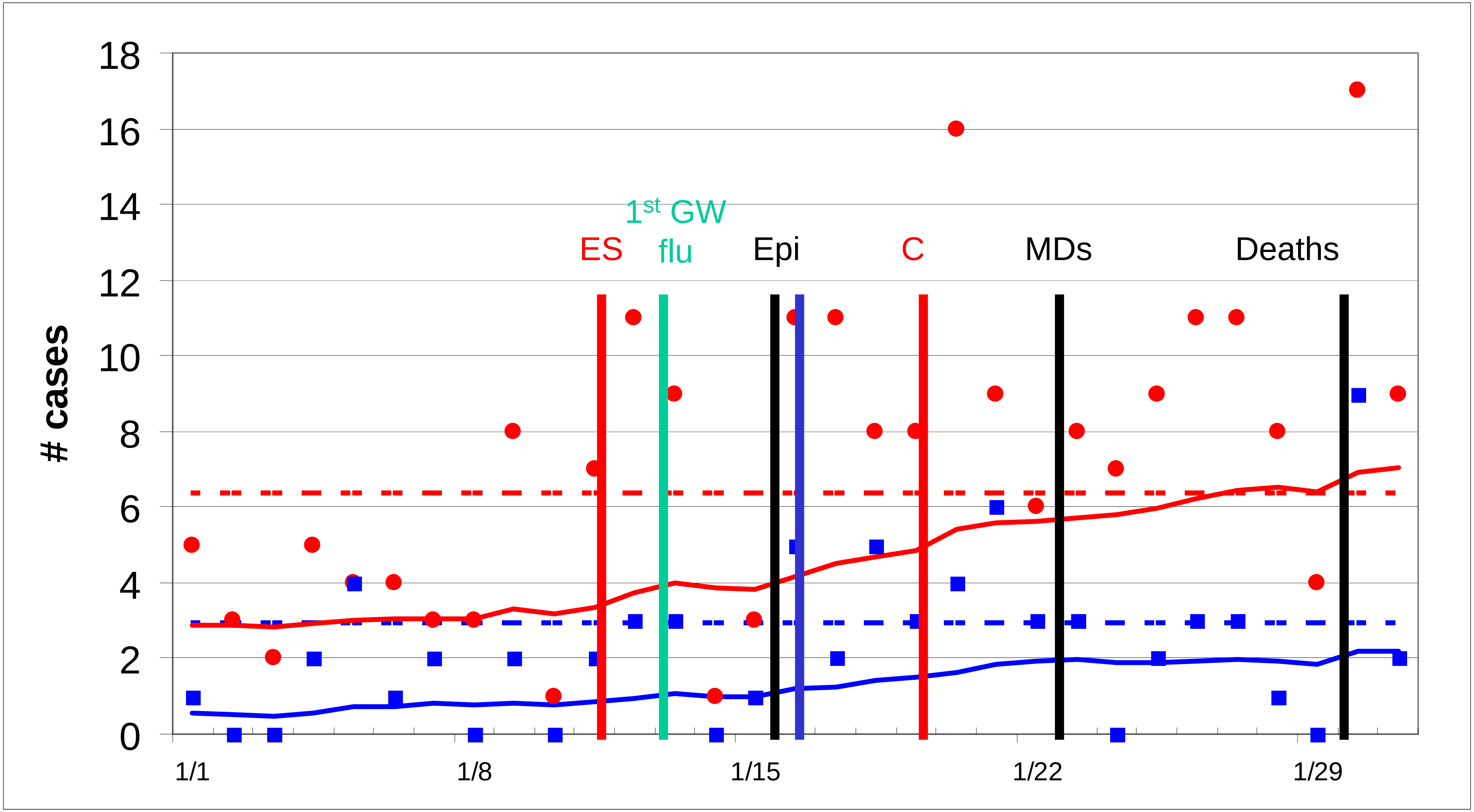


Constant vs. exponential smoothing CUSUM for Viral NOS

- Set k=0.5 for ExSm and k= 1 for constant, and h=2, all years
- First date of ILI outbreak

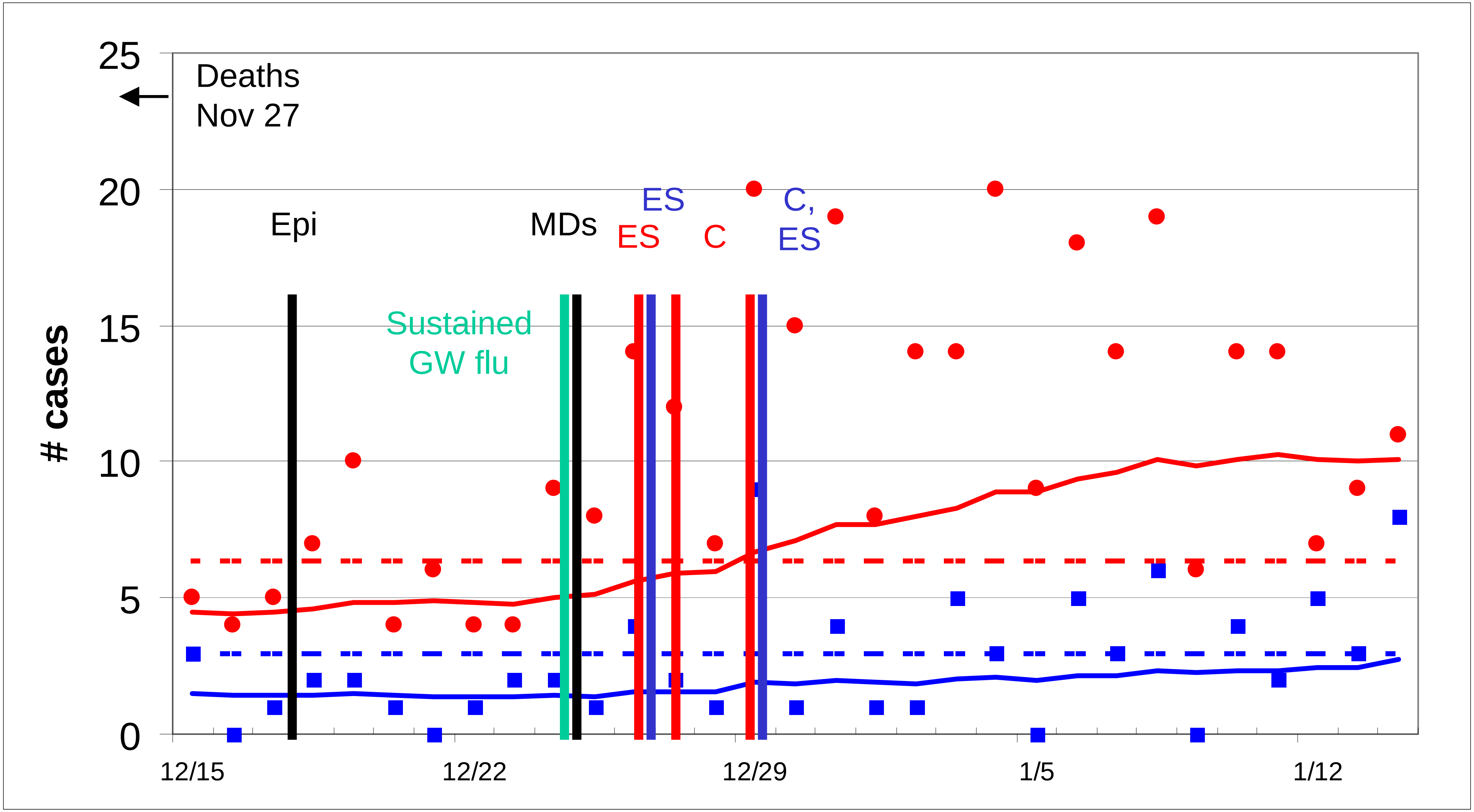
Year	Start of "outbreak"				
	CUSUM		Regional/ widespread	Sentinal MDs	P&I Deaths
	Constant	Exp. Sm.			
1999	Jan 16	Jan 16	Jan 16	Jan 23	Jan 30
2000	Dec 29	Dec 26/29	Dec 18	Dec 25	Nov 27
2001	Nov 19/Jan 4	Nov 19/Jan 4	Nov 18/Jan 6	Jan 20	X

ILI and Viral NOS, January 1999



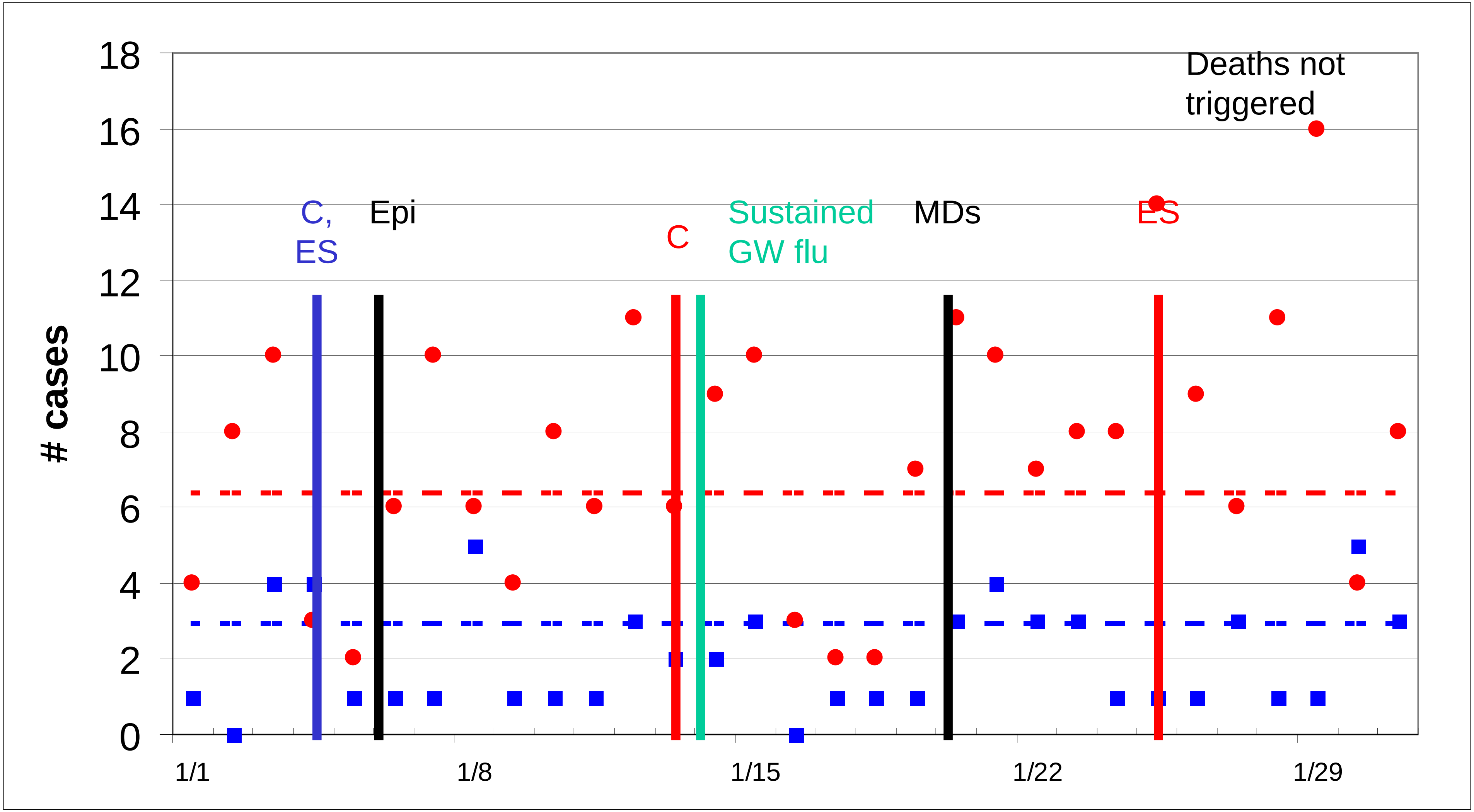
ILI counts, exp. sm., upper CI   Viral NOS counts, exp. sm., upper CI

ILI and Viral NOS, December 1999 - January 2000



ILI counts, exp. sm., upper CI   Viral NOS counts, exp. sm., upper CI

ILI and Viral NOS, January 2001



ILI counts, exp. sm., upper CI   Viral NOS counts, exp. sm., upper CI

Summary and conclusions

Start of outbreaks are hard to detect!

- especially when onset is not sharp

For ILI and Viral NOS, CUSUM methods are preferable to standard Z-score methods

- Fewer false positives and no delay in recognizing outbreaks

Neither exponential smoothing nor constant (Poisson) CUSUM methods are clearly preferable to the other

Detection of a bioterrorist event

Depends on the season

- **Fewer cases needed in summer than in winter to arose suspicion**

Influenza diagnosis not usually made in ER outside of flu season, so focus on ILI and Viral NOS

CUSUM methods may be more effective as an automatic detection method for bioterrorism than for the flu season

Additional research

Simultaneous analysis of multiple data streams

- different symptom and ICD code groupings
- different hospitals and service settings
- non-health care data, such as over-the-counter flu medications or absenteeism

Geographical and socioeconomic patterns

As background for these analyses, need to more fully understand the relationship among these indicators in “regular” flu seasons

Run simulations in which public health officials must decide when to “call” an outbreak