What's wrong with evaluating "syndromic surveillance"?

Steven C. Macdonald PhD, MPH
Washington State Dept. of Health
Office of Epidemiology

Evaluation using a “Gold standard”

		Gold Standard: **Yes**	Gold Standard: **No**
Method under evaluation	**Yes**	True positive	False positive
	No	False negative	

Sensitivity & Predictive Value Positive

- Sensitivity – Focus on False Negatives (FN)
  - What cases were not found?
  - Defined as “probability that any given case will be identified”
  - Calculated as Sens = TP / (TP + FN)
- PVP – Focus on False Positives (FP)
  - Can we trust the data we have?
  - Defined as “probability that a case which is labeled positive is truly positive”
  - Calculated as PVP = TP / (TP + FP)

If the prevalence is Zero

		True condition: **Yes**	True condition: **No**
Method under evaluation	**Yes**	0	False positive
	No	0	True negative

Simulation 1

- Assume
  - Prevalence = 1 in 1,000,000
  - Sensitivity = 100%
  - Specificity = 99%
  - 1,000,000 events are screened
- Results
  - The one true event will be found
  - PVP = 0.01% [1/(1+9999.9)]
  - 10,000 FPs for every TP
  - PVN = 100% [there are no FN]

Simulation 2

- Assume
  - Prevalence = 1 in 1,000,000
  - Sensitivity = 85%
  - Specificity = 85%
- Results
  - PVP = 0.0006% [0.85/(0.85+149999.9)]
  - Over 176,000 FPs for every TP
  - PVN = ~100%

Lessons from simulations

- When prevalence is extremely low
  - PVP approaches zero
  - PVN approaches 100%
- Efforts to improve sensitivity & specificity can increase PVP
  - But, as long as prevalence is low, the ratio of FP to TP will be high
- Standard quantitative evaluation of syndromic surveillance will always show unacceptable ratio of FP to TP
  - PVP is virtually zero
  - high burden of wasted time investigating cases which are virtually never true events
- PVN can be falsely reassuring, if viewed in isolation
- If system has never experienced a true event, the results are known in advance

Evaluation alternatives - 1

- Process evaluation
  - Focus on Sens/PVP of case ascertainment for syndrome, not target disease
- Problem – does not tell you what you really want to know: How good is the system at finding cases of the target disease?

Evaluation alternatives - 2

- Qualitative attributes
  - Simplicity
  - Flexibility
  - Data quality (for each included case)
  - Acceptability
  - Representativeness (person, place, time)
  - Timeliness
  - Stability
- Compare each attribute for syndromic vs. traditional case-finding
- Problem – does not tell you what you really want to know: How good is the system at finding cases of the target disease?

Evaluation alternatives - 3

- Simulations – Retrospective
  - Example – 1993 Milwaukee *Cryptosporidium* outbreak
    - Review of ED records: Unrecognized increase in diarrhea preceded outbreak notification
  - "Texas sharpshooter" fallacy – Drawing the circle on the barn after shooting the rifle
- Simulations – Prospective
  - Inserting simulated outbreaks into existing surveillance data streams
    - to test analytic methods
    - to test the response of investigators
  - Disruptive of normal operations (requires deception)

So, what's wrong?

- Standard quantitative evaluation gives answers we don't want
  - when we focus on detection of the target disease
- Alternatives are somewhat unsatisfactory
- Evaluation remains a challenge