NAVAL
POSTGRADUATE
SCHOOL

Optimizing a System of Threshold-based Sensors

with Application to Biosurveillance

Ronald D. Fricker, Jr.

Third Annual Quantitative Methods in
Defense and National Security Conference

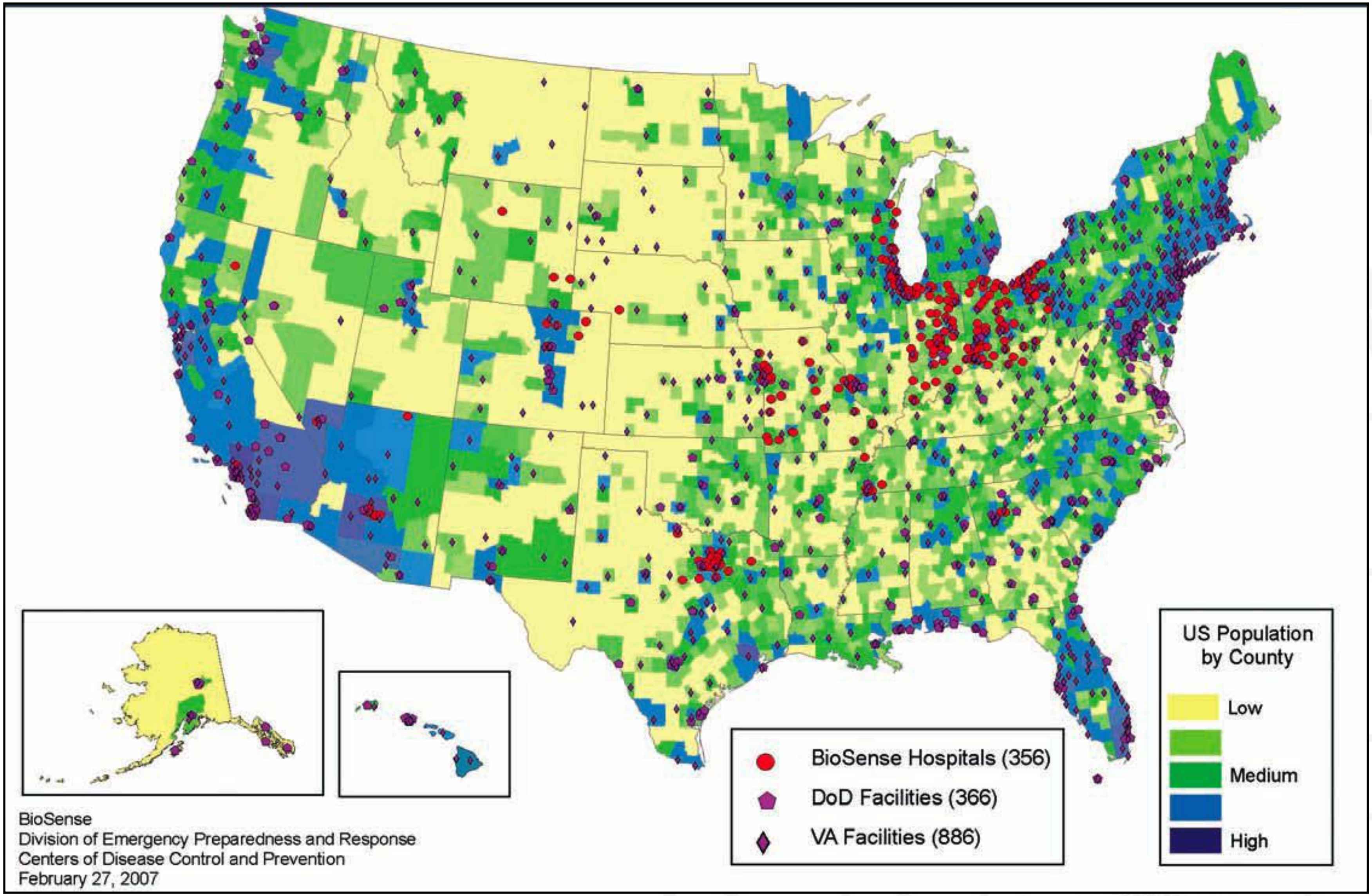
May 28, 2008

What is Biosurveillance?

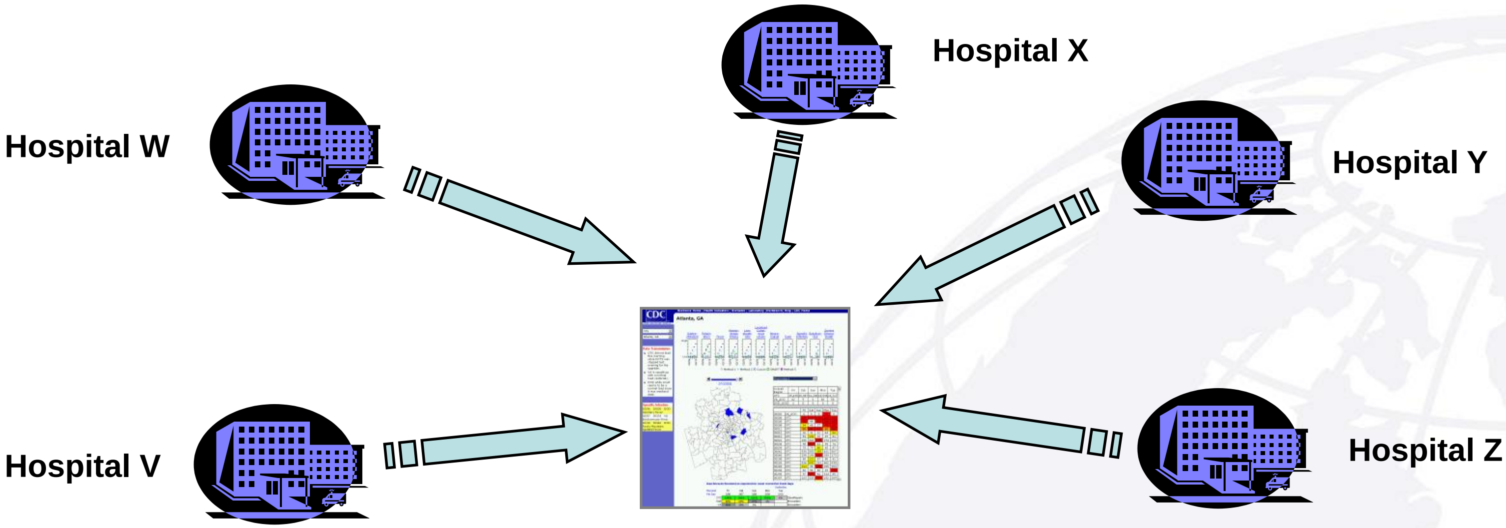
- Homeland Security Presidential Directive HSPD-21 (October 18, 2007):
  - "The term 'biosurveillance' means the process of active data-gathering … of biosphere data … in order to achieve early warning of health threats, early detection of health events, and overall situational awareness of disease activity." [1]
  - "The Secretary of Health and Human Services shall establish an operational national epidemiologic surveillance system for human health..." [1]
- Epidemiologic surveillance:
  - "…surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response." [2]

[1] www.whitehouse.gov/news/releases/2007/10/20071018-10.html
[2] CDC (www.cdc.gov/epo/dphsi/syndromic.htm, accessed 5/29/07)

An Existing System: BioSense



Think of It Like a Large System of Sensors



- Issue: False alarms a serious problem
  - “…*most health monitors… learned to ignore alarms triggered by their system. This is due to the excessive false alarm rate that is typical of most systems - there is nearly an alarm every day!*” [1]

[1] https://wiki.cirg.washington.edu/pub/bin/view/Isds/SurveillanceSystemsInPractice

The Problem in Summary

- Goal: Early detection of disease outbreak and/or bioterrorism
- Issue: Currently detection thresholds set naively
  - Equally for all sensors
  - Ignores differential probability of attack
- Result:
  - High false alarm rates
  - Loss of credibility

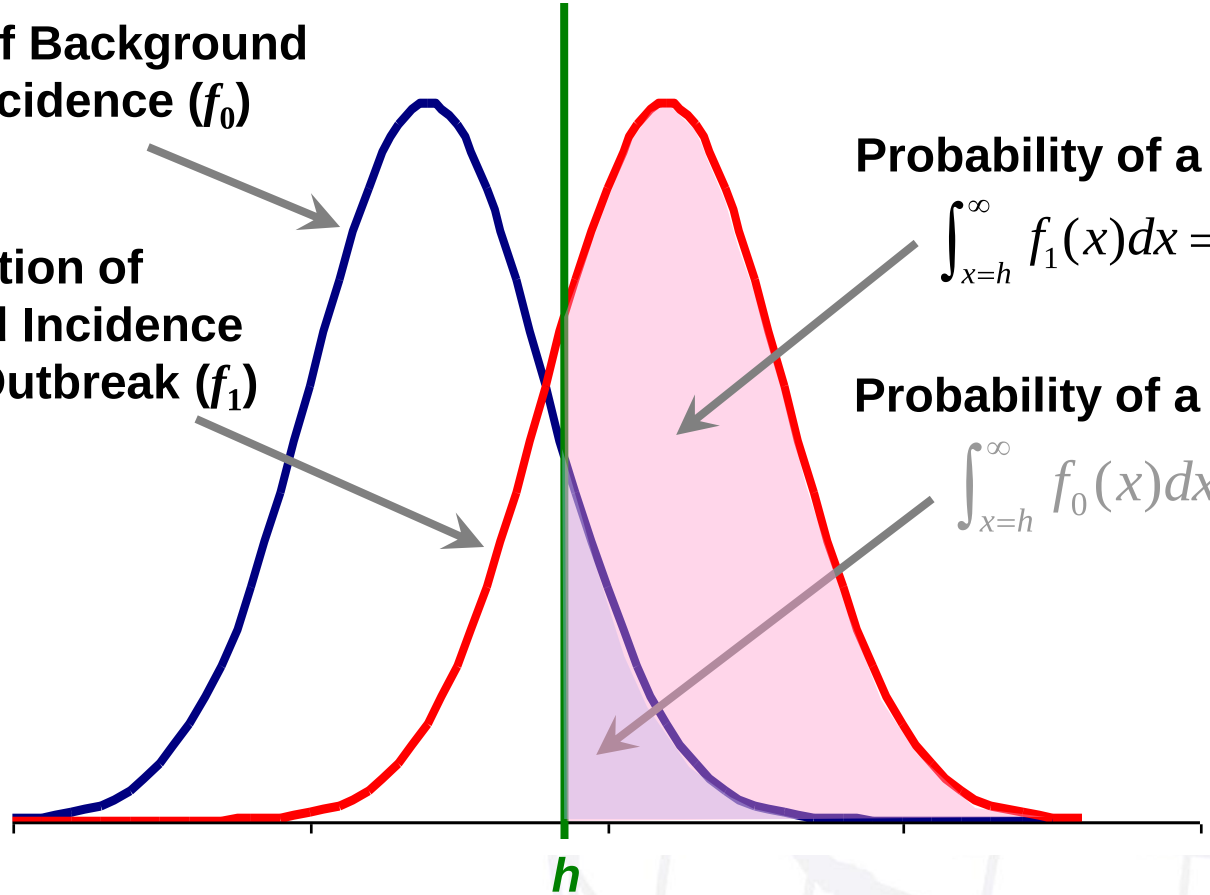
Formal Description of the System

- Let X_{it} denote the output from sensor i at time t, i=1,…,n, t=1,2,…
  - Each sensor / location has a probability of outbreak / attack: $p_1,\ldots,p_n,\ \sum_i p_i = 1$
  - If no "event of interest" anywhere in the network, $X_{it}\sim F_0$ for all i and t
  - If an event of interest occurs at time τ, $X_{i\tau}\sim F_1$ for exactly one i
- A signal is generated at time τ^* when $X_{i\tau^*} \geq h_i$ for one or more i

Idea of Threshold Detection

Distribution of Background Disease Incidence (f_0)

Distribution of Background Incidence and Attack/Outbreak (f_1)



Probability of a true signal:

$$\int_{x=h}^{\infty} f_1(x)dx = 1 - F_1(h)$$

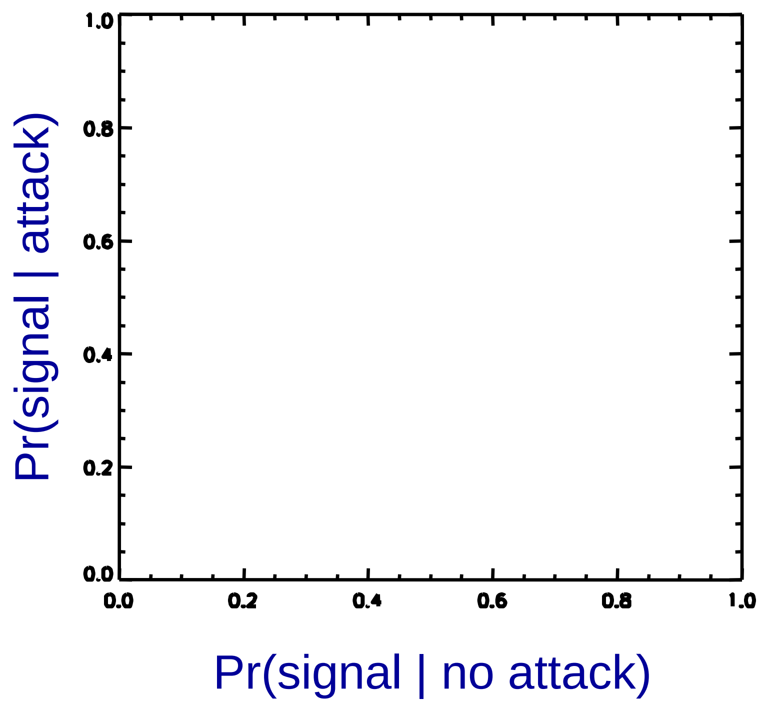
Probability of a false signal:

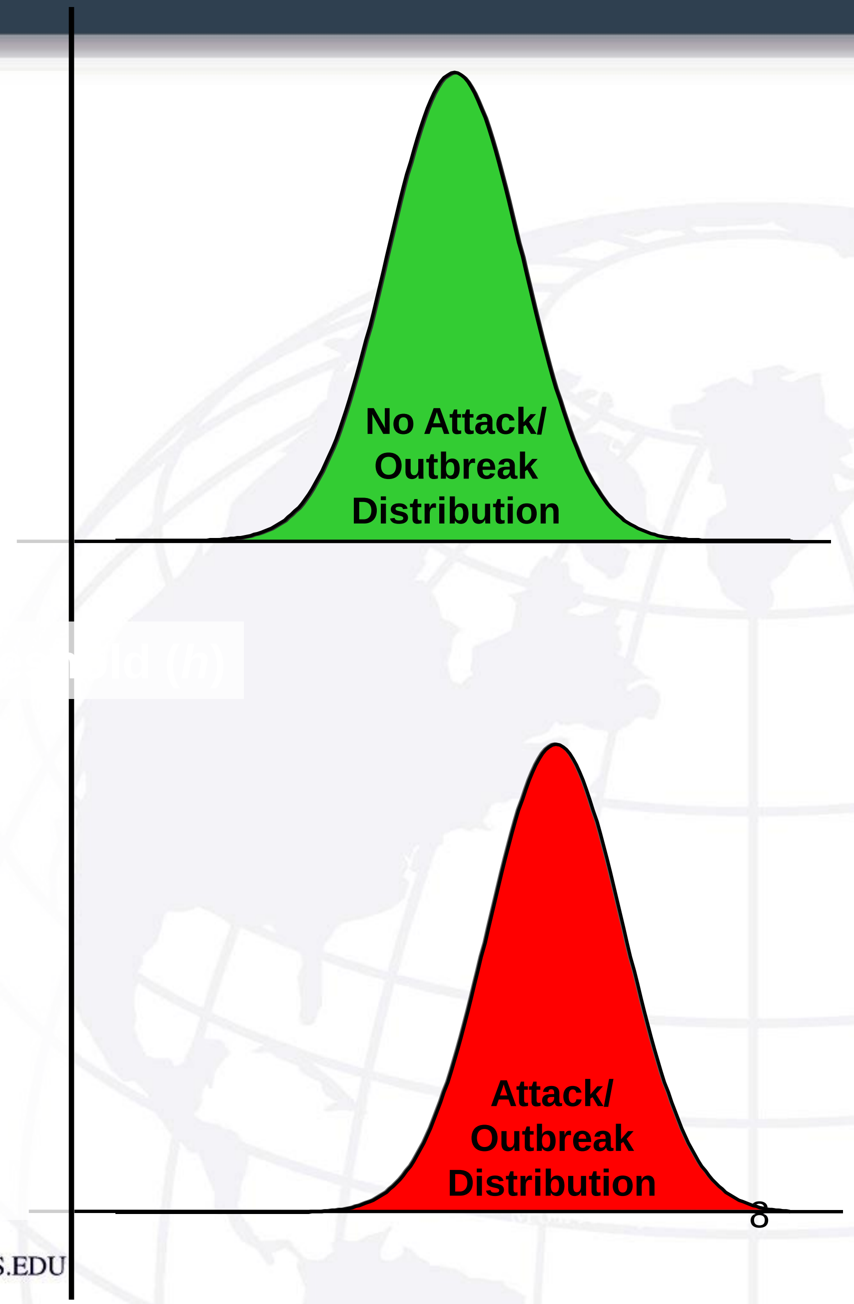
$$\int_{x=h}^{\infty} f_0(x)dx = 1 - F_0(h)$$

It’s All About Choosing Thresholds

- For each sensor, choice of h is compromise between probability of true and false signals

ROC Curve





Mathematical Formulation of the Problem

- It's simple to write out:

$$\Pr(\text{detection}) = \sum_i \Pr(\text{signal}|\text{attack}) \Pr(\text{attack})$$

$$\mathrm{E}(\#\text{ false signals}) = \sum_i \Pr(\text{signal}|\text{no attack})$$

- Express it as an NLP optimization problem:

$$\max_{\vec{h}} \sum_i [1 - F_1(h_i)] p_i$$

$$\text{s.t.} \sum_i [1 - F_0(h_i)] \le \kappa$$

Some Assumptions

- Sensors are spatially independent
- Monitoring standardized residuals from an "adaptive regression" model
  - Model accounts for (and removes) systematic effects in the data
  - Result: Reasonable to assume F_0=N(0,1)
- An attack will result in a 2-sigma increase in the mean of the residuals
  - Result: F_1=N(2,1)
- Then, NLP is:

$$\min_{\vec{h}} \sum_i \Phi(h_i - 2) p_i$$

$$\text{s.t.} \sum_i \Phi(h_i) > n - \kappa$$

Ten Sensor Example

Sensor i	p_i	Common Threshold #1	Optimal Threshold (h_i)	Common Threshold #2
1	0.797	2.189	1.068	1.310
2	0.064	2.189	3.602	1.310
3	0.056	2.189	3.732	1.310
4	0.048	2.189	3.915	1.310
5	0.013	2.189	4.656	1.310
6	0.006	2.189	4.736	1.310
7	0.006	2.189	4.736	1.310
8	0.005	2.189	4.755	1.310
9	0.003	2.189	4.773	1.310
10	0.002	2.189	4.791	1.310
	P_d	0.117	0.378	0.378
	$\sum \alpha_i$	0.143	0.143	0.951

NAVAL POSTGRADUATE SCHOOL

Simplifying to a One-dimensional Optimization Problem

- System of n hospitals (sensors) means optimization has n free parameters
  - Hard for to solve for large systems
- Can simplify to one-parameter problem:
  - *Theorem*: For F_0=N(0,1) and F_1=N(γ,1), the optimization simplifies to finding μ to satisfy

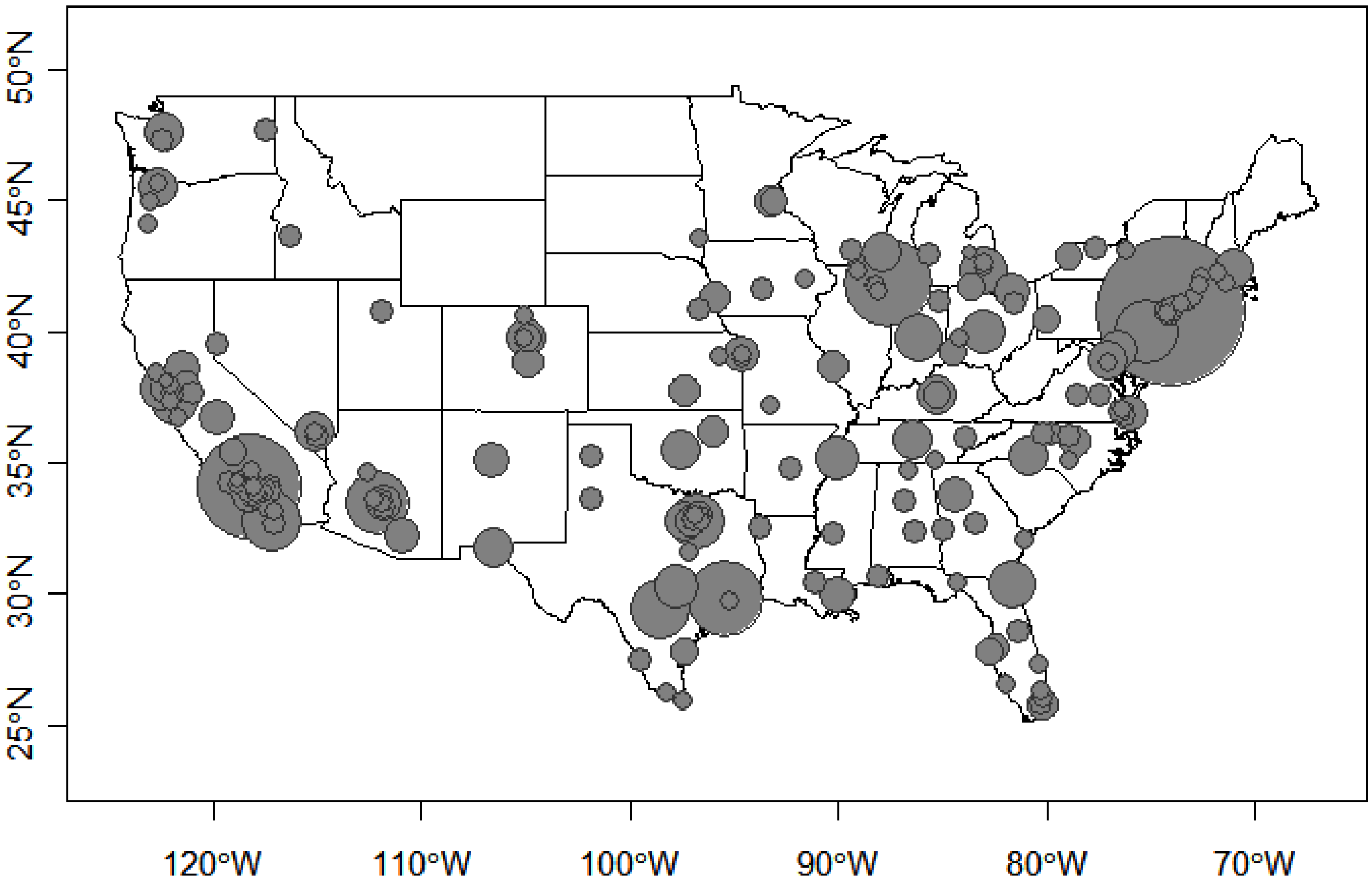
$$\sum_{i=1}^{n} \Phi\left(\mu - \frac{1}{\gamma} \ln(p_i) \right) = n - \kappa,$$

and the optimal thresholds are then

$$h_i = \mu - \frac{1}{\gamma} \ln(p_i).$$

Consider (Hypothetical) System to Monitor 200 Largest Cities in US

- Assume probability of attack is proportional to the population in a city: $p_i = m_i / \sum_i m_i$



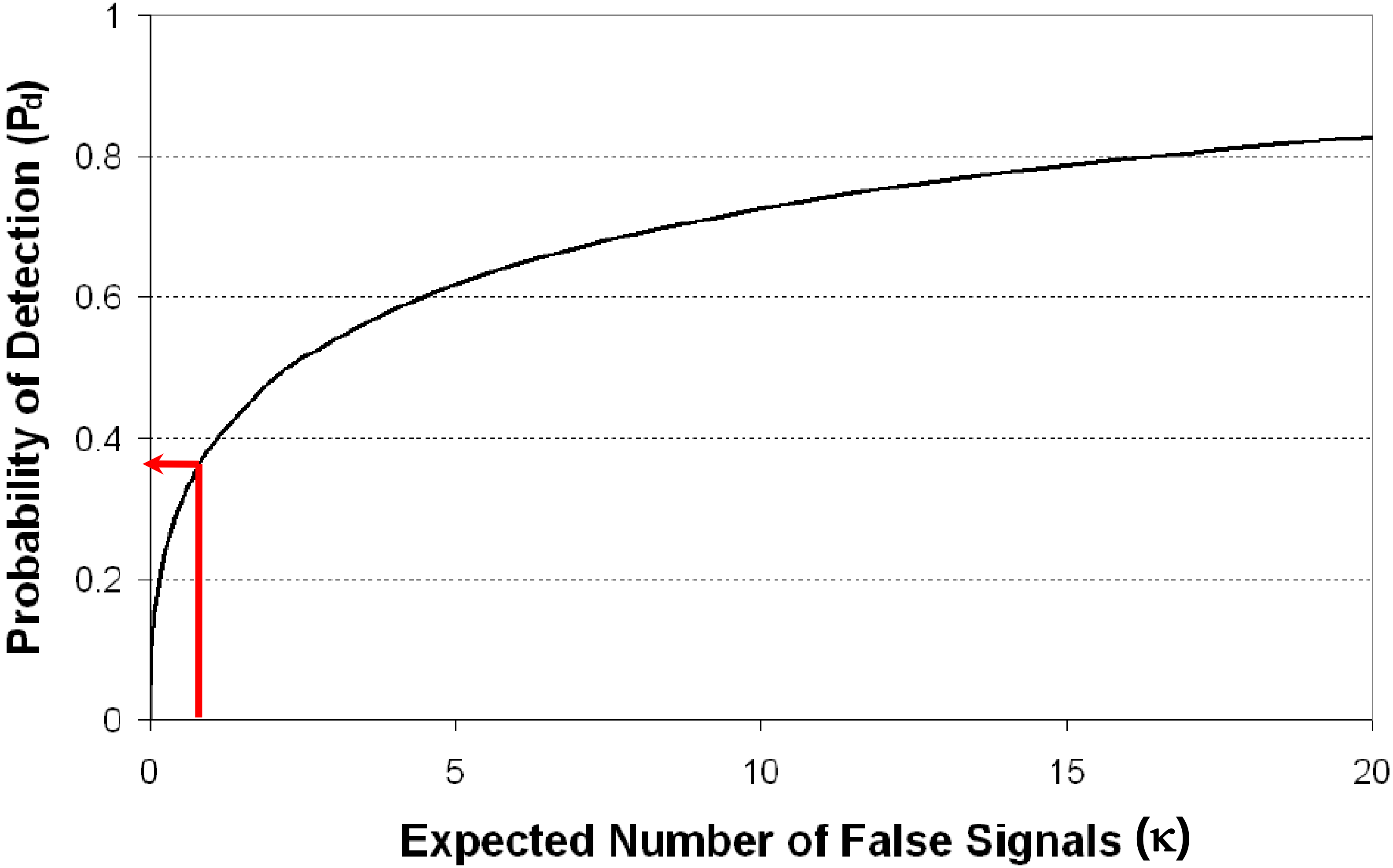
Optimal Solution for 200 Cities

- Assume
  - *2σ* magnitude event
  - Constraint of 1 false signal system-wide / day

				Population	Pr(attack)	Threshold	Pr(signal | attack)	Pr(signal | no attack)
7	i	City	State	m_i	$p_i = m_i / M$	h_i	$1-\Phi(h_i-\delta)$	$1-\Phi(h_i)$
8	1	New York city	New York	8,214,426	0.1101	1.07	0.825	0.143
9	2	Los Angeles	California	3,849,378	0.0516	1.45	0.710	0.074
10	3	Chicago	Illinois	2,833,321	0.0380	1.60	0.656	0.055
11	4	Houston	Texas	2,144,491	0.0287	1.74	0.603	0.041
12	5	Phoenix	Arizona	1,512,986	0.0203	1.91	0.535	0.028
13	6	Philadelphia	Pennsylvania	1,448,394	0.0194	1.93	0.526	0.027
14	7	San Antonio	Texas	1,296,682	0.0174	1.99	0.504	0.023
15	8	San Diego	California	1,256,951	0.0168	2.01	0.498	0.022
16	9	Dallas	Texas	1,232,940	0.0165	2.01	0.494	0.022
17	10	San Jose	California	929,936	0.0125	2.16	0.438	0.016

- Result: Pr(signal | attack) = 0.388
- Naïve result: Pr(signal | attack) = 0.283

P_d – False Alarm Trade-Off



Choosing γ and κ

- Optimal probability of detection for various choices of γ and κ

$\mathbf{P}_d$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
$\gamma = 1$	0.165	0.228	0.272	0.307	0.336
$\gamma = 2$	0.388	0.481	0.540	0.583	0.618
$\gamma = 3$	0.726	0.801	0.840	0.866	0.885
$\gamma = 4$	0.939	0.964	0.974	0.980	0.984

  – Choice of κ depends on available resources
  – Setting γ is subjective: what size mean increase important to detect?

Sensitivity Analyses

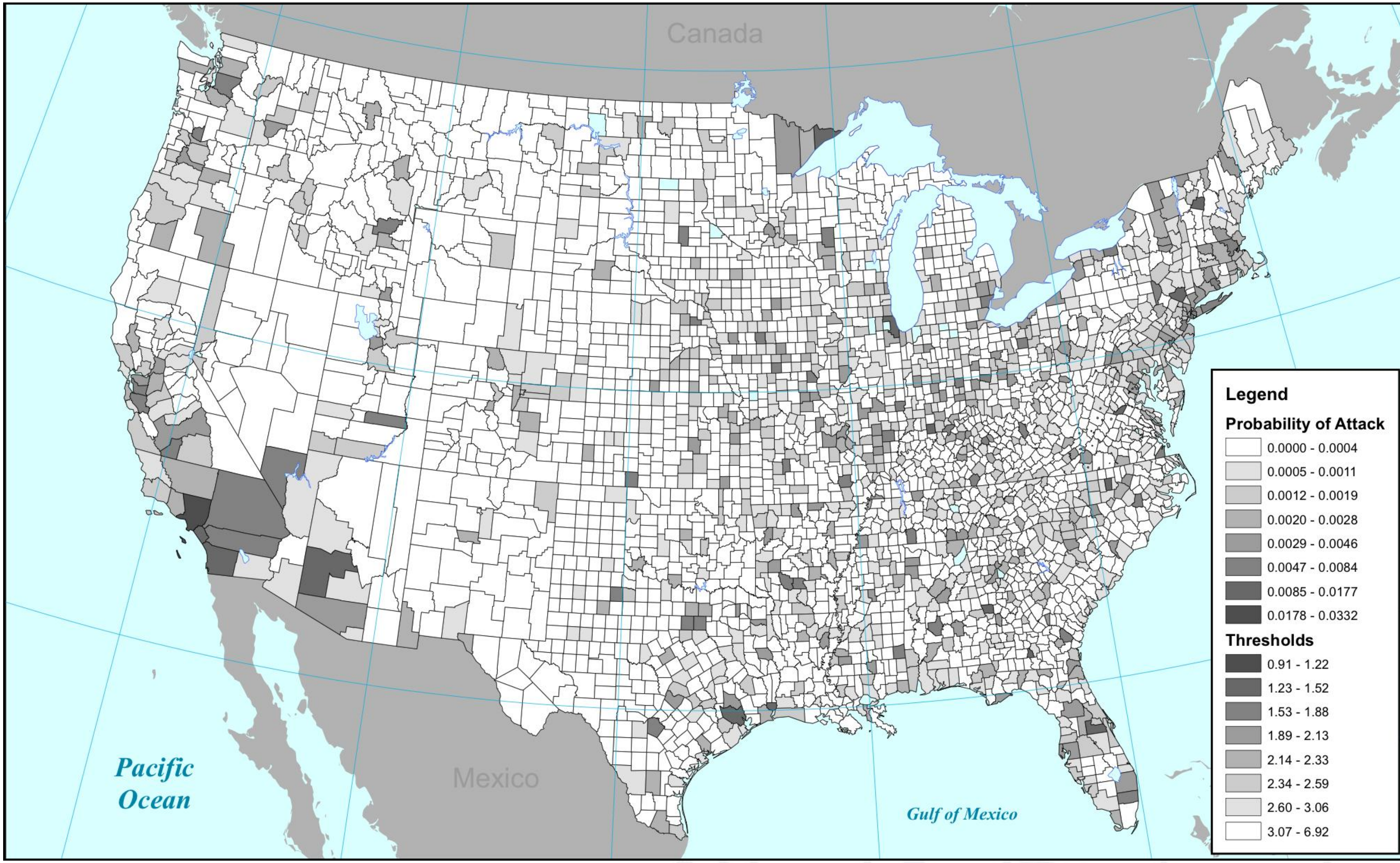
- Optimal probability of detection

$\mathbf{P}_d$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
$\gamma = 1$	0.165	0.228	0.272	0.307	0.336
$\gamma = 2$	0.388	0.481	0.540	0.583	0.618
$\gamma = 3$	0.726	0.801	0.840	0.866	0.885
$\gamma = 4$	0.939	0.964	0.974	0.980	0.984

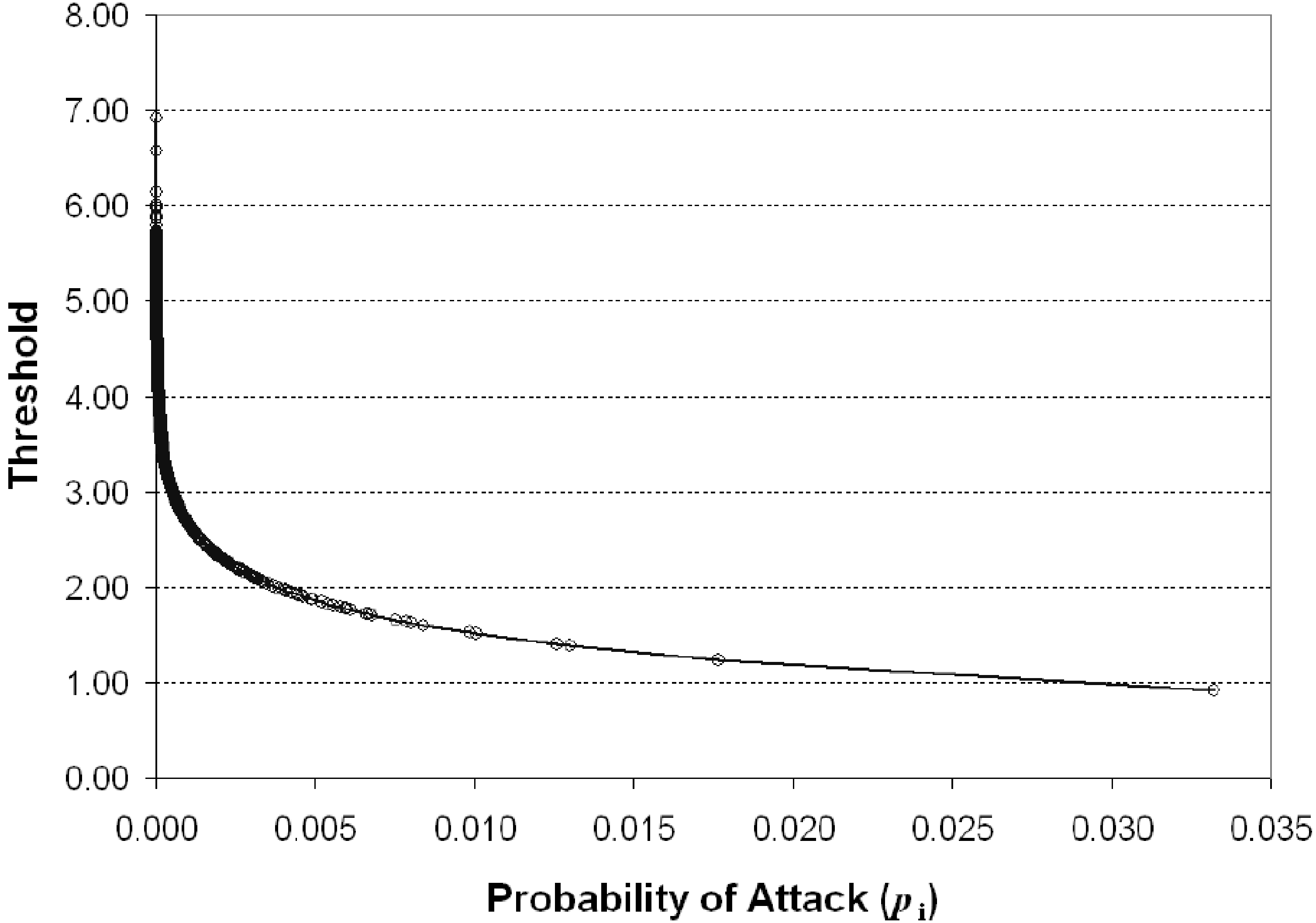
- Actual probability of detection

$\mathbf{P}_d$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
Observed $\gamma = 1$	0.137	0.193	0.235	0.269	0.298
Observed $\gamma = 2$	0.388	0.481	0.540	0.583	0.618
Observed $\gamma = 3$	0.711	0.790	0.832	0.859	0.879
Observed $\gamma = 4$	0.925	0.955	0.968	0.976	0.981

Optimizing a County-level System



Thresholds as a Function of Probability of Attack



Take-Aways

- BioSense and other biosurveillance systems' performance can be improved now at no cost
- Approach allows for customization
  - E.g., increase in probability of detection at individual location or add additional constraint to minimize false signals
- Applies to other sensor system applications:
  - Port surveillance, radiation/chem detection systems, etc.
- Details in Fricker and Banschbach (2007)

Future Research Directions

- Assess data fusion techniques for use when multiple sensors in each region
  - I.e., relax sensor (spatial) independence assumption
- Generalize from threshold detection methods to other methods that use historical information
  - I.e., relax temporal independence assumption

Selected References

Background Information:

- Fricker, R.D., Jr., and H. Rolka, Protecting Against Biological Terrorism: Statistical Issues in Electronic Biosurveillance, *Chance*, **91**, pp. 4-13, 2006
- Fricker, R.D., Jr., Syndromic Surveillance, *Encyclopedia of Quantitative Risk Assessment* (to appear).

Selected Research:

- Fricker, R.D., Jr., and D. Banschbach, Optimizing a System of Threshold Detection Sensors, in submission to *Operations Research*.
- Fricker, R.D., Jr., and J.T. Chang, A Spatio-temporal Method for Real-time Biosurveillance, *Quality Engineering*, (to appear).
- Fricker, R.D., Jr., Knitt, M.C., and C.X. Hu, Comparing Directionally Sensitive MCUSUM and MEWMA Procedures with Application to Biosurveillance, *Quality Engineering* (to appear).
- Joner, M.D., Jr., Woodall, W.H., Reynolds, M.R., Jr., and R.D. Fricker, Jr., A One-Sided MEWMA Chart for Health Surveillance, *Quality and Reliability Engineering International* (to appear).
- Fricker, R.D., Jr., Hegler, B.L., and D.A Dunfee, Assessing the Performance of the Early Aberration Reporting System (EARS) Syndromic Surveillance Algorithms, *Statistics in Medicine,* 2008.
- Fricker, R.D., Jr., Directionally Sensitive Multivariate Statistical Process Control Methods with Application to Syndromic Surveillance, *Advances in Disease Surveillance,* **3**:1, 2007.