Evaluation of Anomalous Data Clusters Found in Electronic Surveillance Systems

Presented at the 2nd Annual PHIN Users Conference
May 27th 2004

Joe Lombardo
Howard Burkom
Julie Casani
Jacki Coberly
Kathy Hurt-Mullen
Julie Pavlin
Denise Sockwell
Don Weiss

3.04.108.PSD

Issues Facing the Surveillance System User:

1) Review of Data: Can a set of protocols be identified for systematic review of data captured by electronic surveillance systems?

2) Investigation of Alerts: Can a set of protocols be identified to systematically investigate flags raised by analytical techniques applied to syndromic data?

3) Support Capability: What functions should an automated system provide to expedite follow-up of analytically generated or user-generated alerts?

Working Group

Public Health

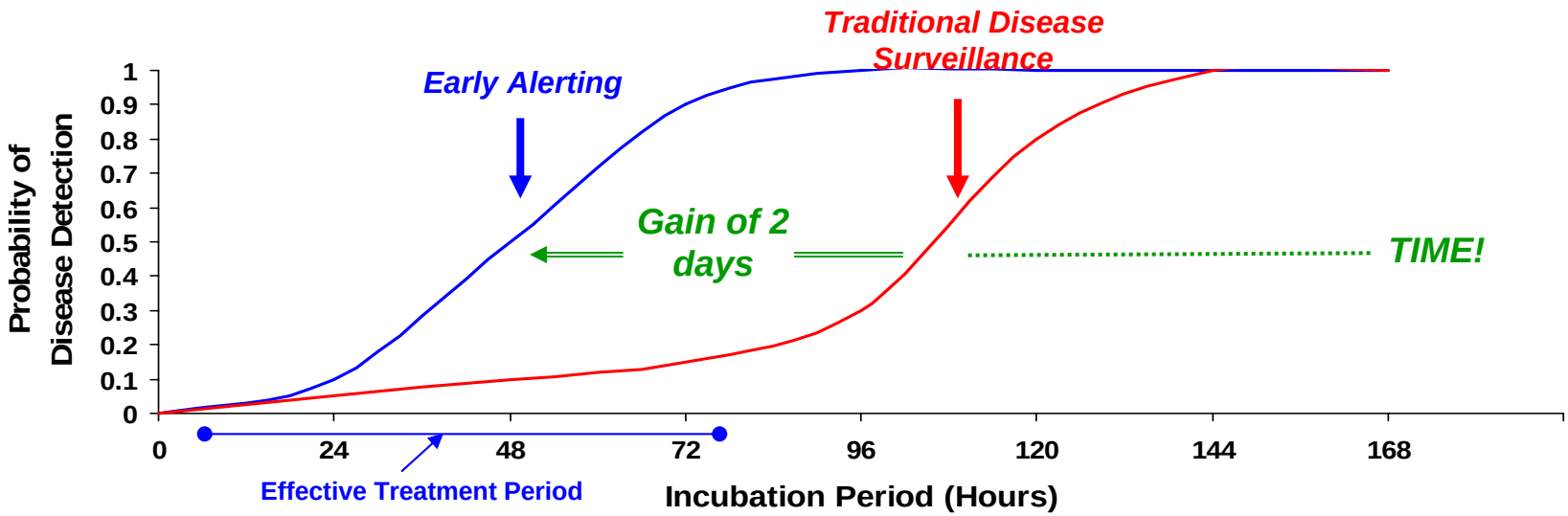
- **Julie Casani** — **Maryland Department of Health and Mental Hygiene**
- **Kathy Hurt-Mullen** — **Montgomery County Health Department**
- **Denise Sockwell** — **Virginia Department of Health**
- **Don Weiss** — **New York City Department of health and Mental Hygiene**

Military Preventive Medicine

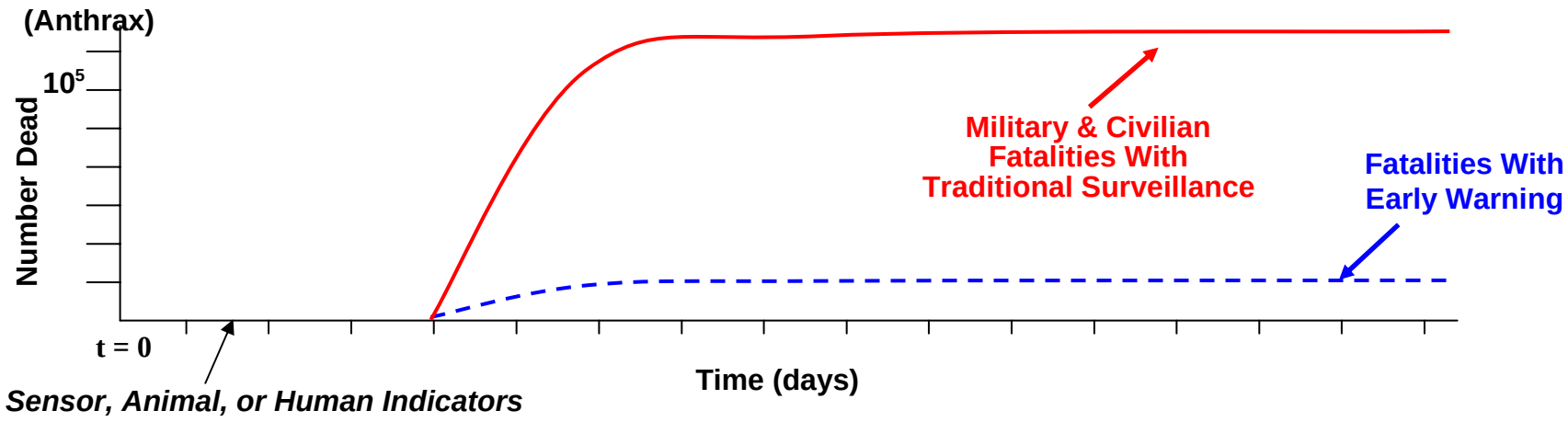
- **Julie Pavlin** — **Walter Reed Army Institute of Research**

- ***Academic & Development***
- **Howard Burkom** — **Johns Hopkins University Applied Physics Laboratory**
- **Jacqueline Coberly** — **Johns Hopkins Bloomberg School of Public Health**
- **Joe Lombardo** — **Johns Hopkins University Applied Physics Laboratory**

Early Alerting Requirement for BioTerrorism

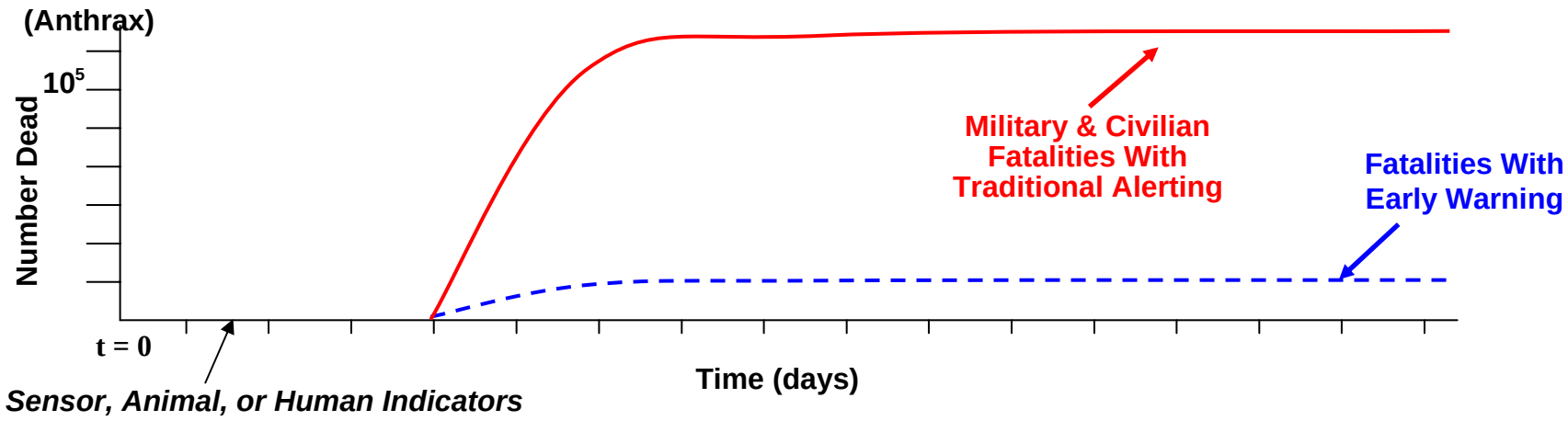


Potential Impact of Early Alerting

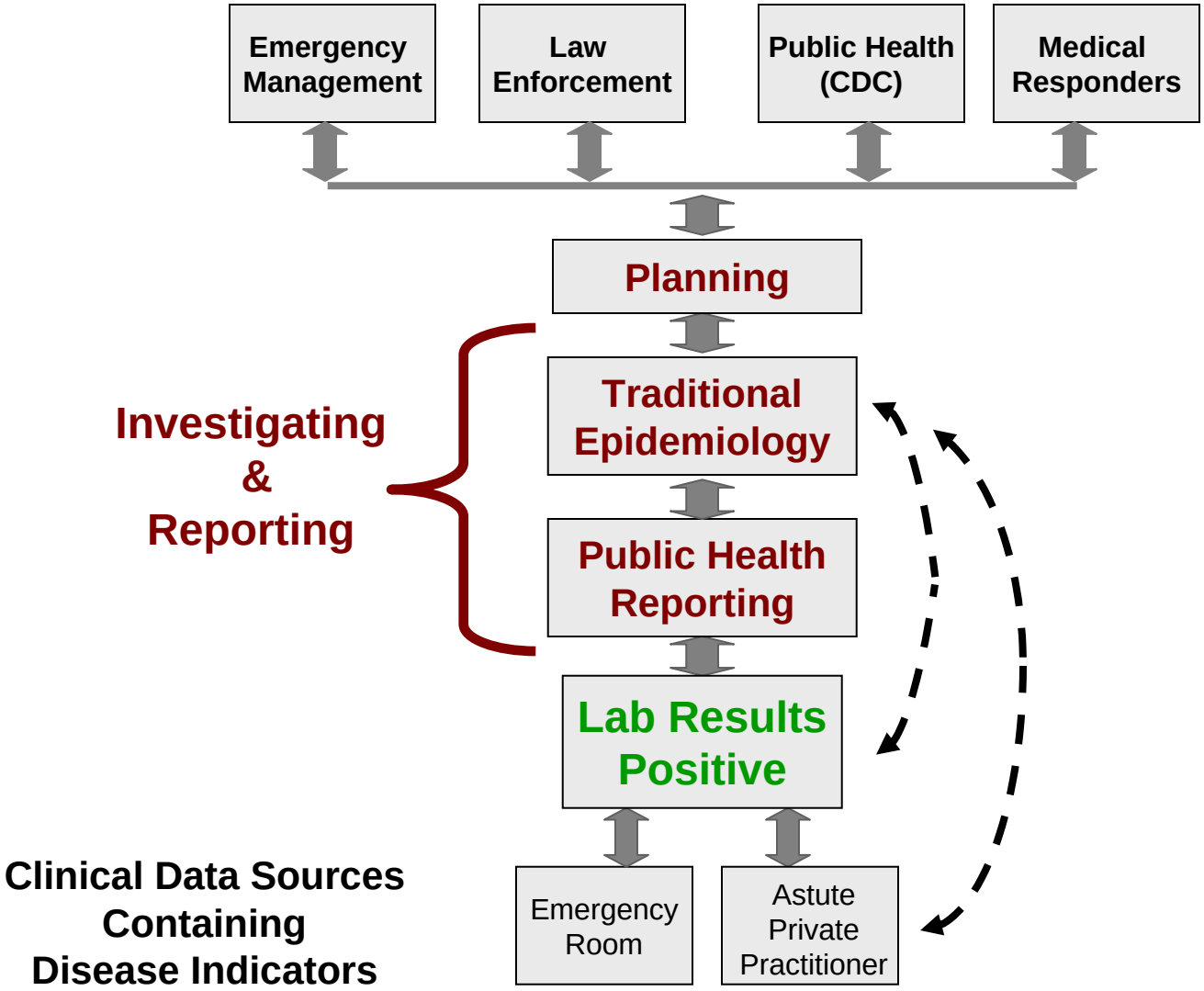


Early Alerting

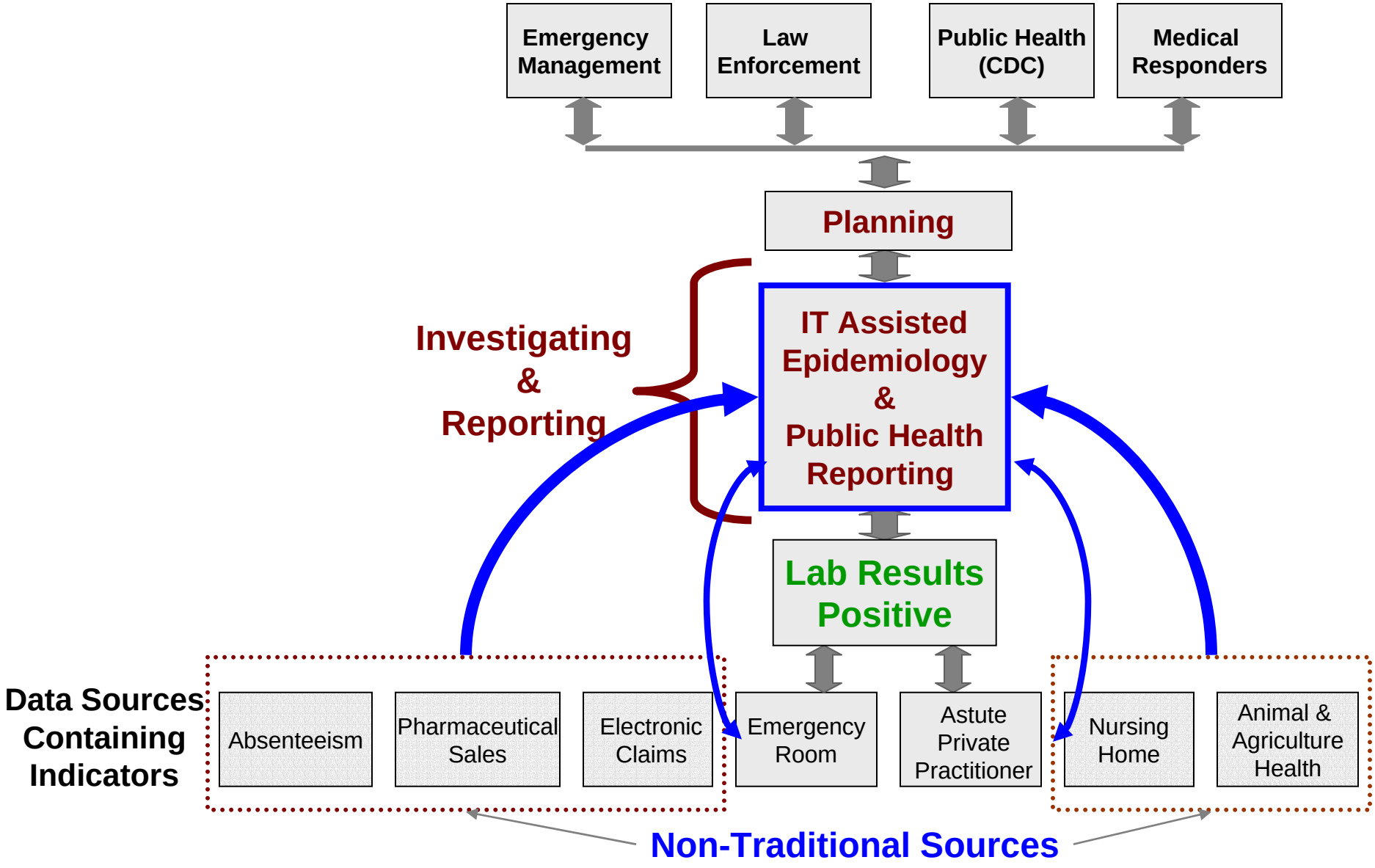
A driving force for the implementation of modern disease surveillance technology is the reduction in mortality and morbidity which could result from a bioterrorist incident.



Traditional Surveillance/Response Process



Enhanced Surveillance/Response Process



Modern Electronic Surveillance

Modern disease surveillance systems have many of the following attributes:

1) Automated data capture, cleansing, archiving, and secure telecommunications

2) Multiple data streams from similar or different providers

3) Extensive archiving and retrieval capacity

4) Outbreak detection algorithms applied to data streams for identification of abnormal conditions, and means for alerting users of these conditions.

5) Web-based distribution of data

6) Geographic representation of data and clusters

Modern Electronic Surveillance

Modern disease surveillance systems have many of the following attributes:

1) Automated data capture, cleansing, archiving, and secure telecommunications

2) Multiple data streams from similar or different providers

3) Extensive archiving and retrieval capacity

4) Outbreak detection algorithms applied to data streams for identification of abnormal conditions, and means for alerting users of these conditions.

5) Web-based distribution of data

6) Geographic representation of data and clusters

Permits the user to employ many analytical tools and view large quantities of data in a limited time

Syndromic Surveillance

Syndromic Surveillance is the collection and analysis of pre-diagnosis information leading to an estimation of the health status of the community

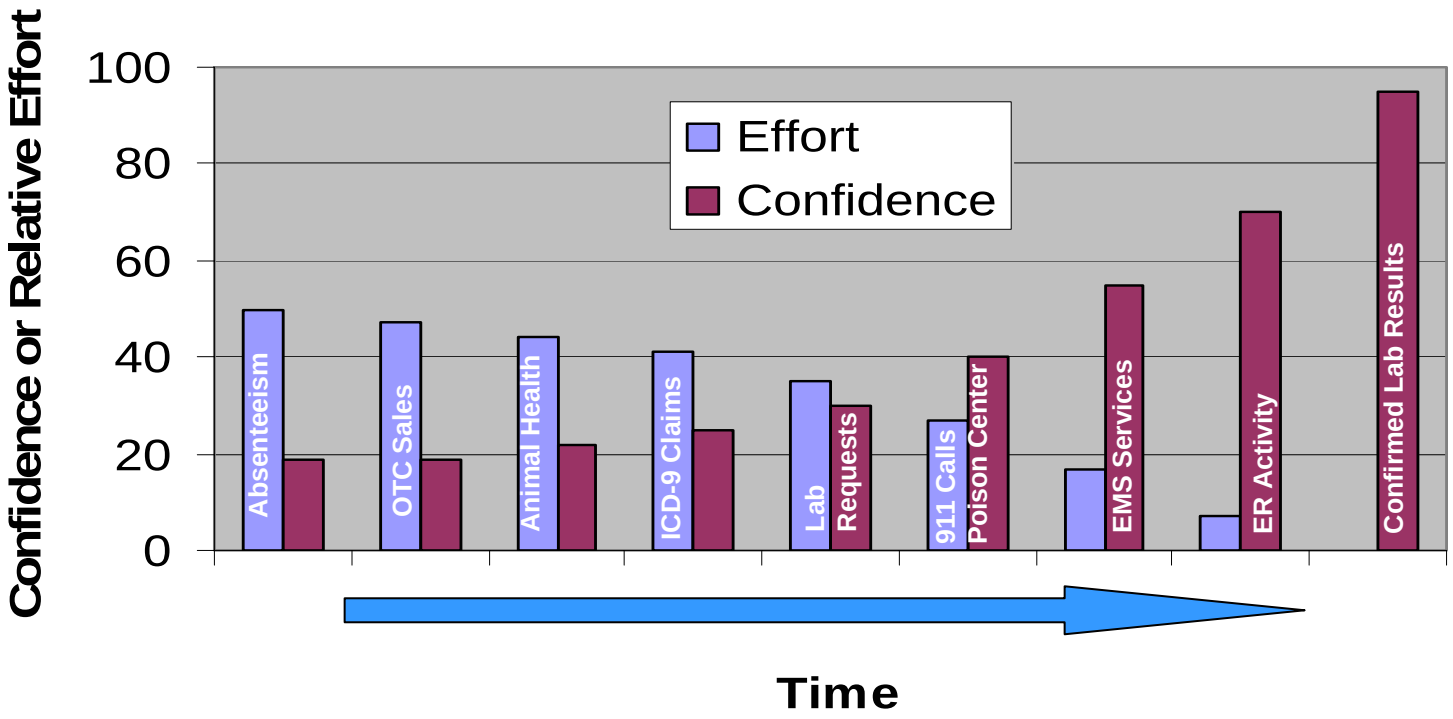
Syndromic surveillance uses health care indicators such as:

- Emergency Room Chief Complaint
- International Classification of Disease Codes
- Requests for Specific Laboratory Tests
- Over the Counter Medications Sales

Grouped into specific syndromic categories such as:

- Respiratory
- Gastrointestinal
- Neurological
- Dermatological
- Fever
- etc.

Impact on Use of Multiple Indicator Sources



Uses for Electronic Surveillance

Electronic Surveillance System

Daily Monitoring

Statistical Anomaly (Alert)

- Characterize w / ESS
  - Age, sex, geography
  - Compare multiple data sources
  - Environmental /other cause
- Request more info from HD/docs/labs, etc
- Public Health Importance
  - Public health significance
  - Intervention warranted?
- Intervention

No Statistical Anomaly

Skepticism requires digging deeper with specialized queries

Data Review

To characterize the health of your community

- Disease patterns by age, sex, time, geography.
- Most common health problems

Uses for Electronic Surveillance

Electronic Surveillance System

Daily Monitoring

Statistical Anomaly (Alert)

- Characterize w / ESS
  - Age, sex, geography
  - Compare multiple data sources
  - Environmental /other cause
- Request more info from HD/docs/labs, etc
- Public Health Importance
  - Public health significance
  - Intervention warranted?
- Intervention

No Statistical Anomaly

Skepticism requires digging deeper with specialized queries

Data Review

To characterize the health of your community

- Disease patterns by age, sex, time, geography.
- Most common health problems

Objective is to acquire knowledge of the population

Uses for Electronic Surveillance

Electronic Surveillance System

Daily Monitoring

Statistical Anomaly (Alert)

- Characterize w / ESS
  - Age, sex, geography
  - Compare multiple data sources
  - Environmental /other cause
- Request more info from HD/docs/labs, etc
- Public Health Importance
  - Public health significance
  - Intervention warranted?
- Intervention

No Statistical Anomaly

Skepticism requires digging deeper with specialized queries

Data Review

To characterize the health of your community

- Disease patterns by age, sex, time, geography.
- Most common health problems

Monitor the abnormal disease levels in the community

Uses for Electronic Surveillance

Electronic Surveillance System

Daily Monitoring

Statistical Anomaly (Alert)

- Characterize w / ESS
  - Age, sex, geography
  - Compare multiple data sources
  - Environmental /other cause
- Request more info from HD/docs/labs, etc
- Public Health Importance
  - Public health significance
  - Intervention warranted?
- Intervention

No Statistical Anomaly

Skepticism requires digging deeper with specialized queries

System is used to address user concerns

Data Review

To characterize the health of your community

- Disease patterns by age, sex, time, geography.
- Most common health problems

User / System Generated Alert

Characterize w / ESS

- age, sex, geography
- compare multiple data sources
- environmental /other cause

Request more info from HD/docs/labs, etc.

Public Health Importance

- Public health significance
- Intervention warranted?

Intervention

Sample Detector Output Listing

History of ESSENCE | Syndrome Definitions | Detector Algorithms | Data Dictionary | Tutorial

HOME | Alert List | Query | Map Portal | Message Board

ESSENCE IV - Demo Alert List

Region/Syndrome Based Temporal Alerts

[Region/Syndrome | Region/ICD9Group | Hospital/Syndrome | Region/OTC Category | Schools | Spatial | Reportable Diseases]

Reset 3-Level Sorting

Region/Syndrome Based Temporal Alerts

Date	Data Source	Region	Age	Sex	Syndrome	Detector	Level	Count	Exp
20Apr04	Emergency Room Data by Patient Location	MONTGOMERY	5-17	All	GastroIntestinal	JHUAPL-Regression/EWMA	2.276	17	7.67
20Apr04	Emergency Room Data by Patient Location	PRINCE GEORGES	18-64	All	Neurological	JHUAPL-Regression/EWMA	4.857	11	2.60
20Apr04	Emergency Room Data by Patient Location	PRINCE GEORGES	All	All	Neurological	JHUAPL-Regression/EWMA	2.016	13	5.53
20Apr04	Emergency Room Data by Patient Location	ARLINGTON	65+	All	Sepsis	JHUAPL-Regression/EWMA	2.033	1	0.35
19Apr04	Emergency Room Data by Patient Location	All	65+	All	Sepsis	JHUAPL-Regression/EWMA	2.231	10	3.10
19Apr04	Emergency Room Data by Patient Location	PRINCE GEORGES	All	All	Respiratory	JHUAPL-Regression/EWMA	2.412	130	80.3
19Apr04	Emergency Room Data by Patient Location	ARLINGTON	65+	All	Sepsis	JHUAPL-Regression/EWMA	2.532	3	0.35
19Apr04	Emergency Room Data by Patient Location	FAIRFAX	65+	All	Sepsis	JHUAPL-Regression/EWMA	3.285	4	0.53
19Apr04	Emergency Room Data by Patient Location	ARLINGTON	All	All	Sepsis	JHUAPL-Regression/EWMA	2.248	3	0.5
19Apr04	Emergency Room Data by Patient Location	FAIRFAX	All	All	Sepsis	JHUAPL-Regression/EWMA	2.303	5	1.35
19Apr04	Emergency Room Data by Patient Location	PRINCE GEORGES	0-4	All	UnspecifiedInfection	JHUAPL-Regression/EWMA	2.785	25	16.6

Internet

Sample Query Function

History of ESSENCE | Syndrome Definitions | Detector Algorithms | Data Dictionary | Tutorial

HOME | Alert List | Query | Map Portal | Message Board

ESSENCE IV - Demo Time Series

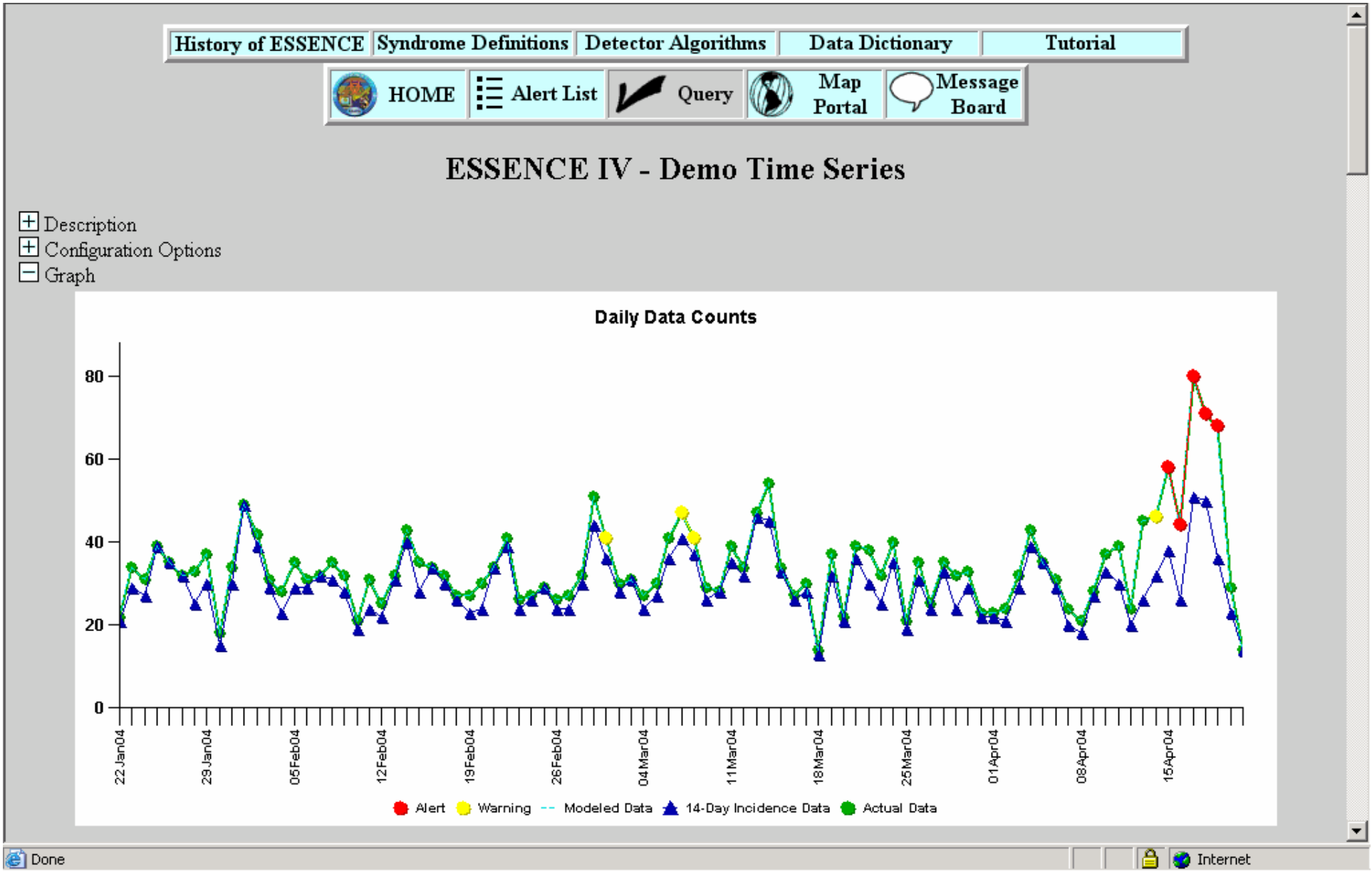
Description
Configuration Options

Data Configuration			
Data Source:	Emergency Room Data by Patient Location	**Region:**	Other_Region Prince Georges, MD Prince William, VA
ESSENCE Syndrome:	Respiratory Sepsis UnspecifiedInfection	**Detector:**	JHUAPL-Regression/EWMA
Age Range:	All Unknown 0-4	**Sex:**	All Unknown Male
Start Date:	22 Jan 04	**End Date:**	21 Apr 04
Change Configuration			

Graph

Done | Internet

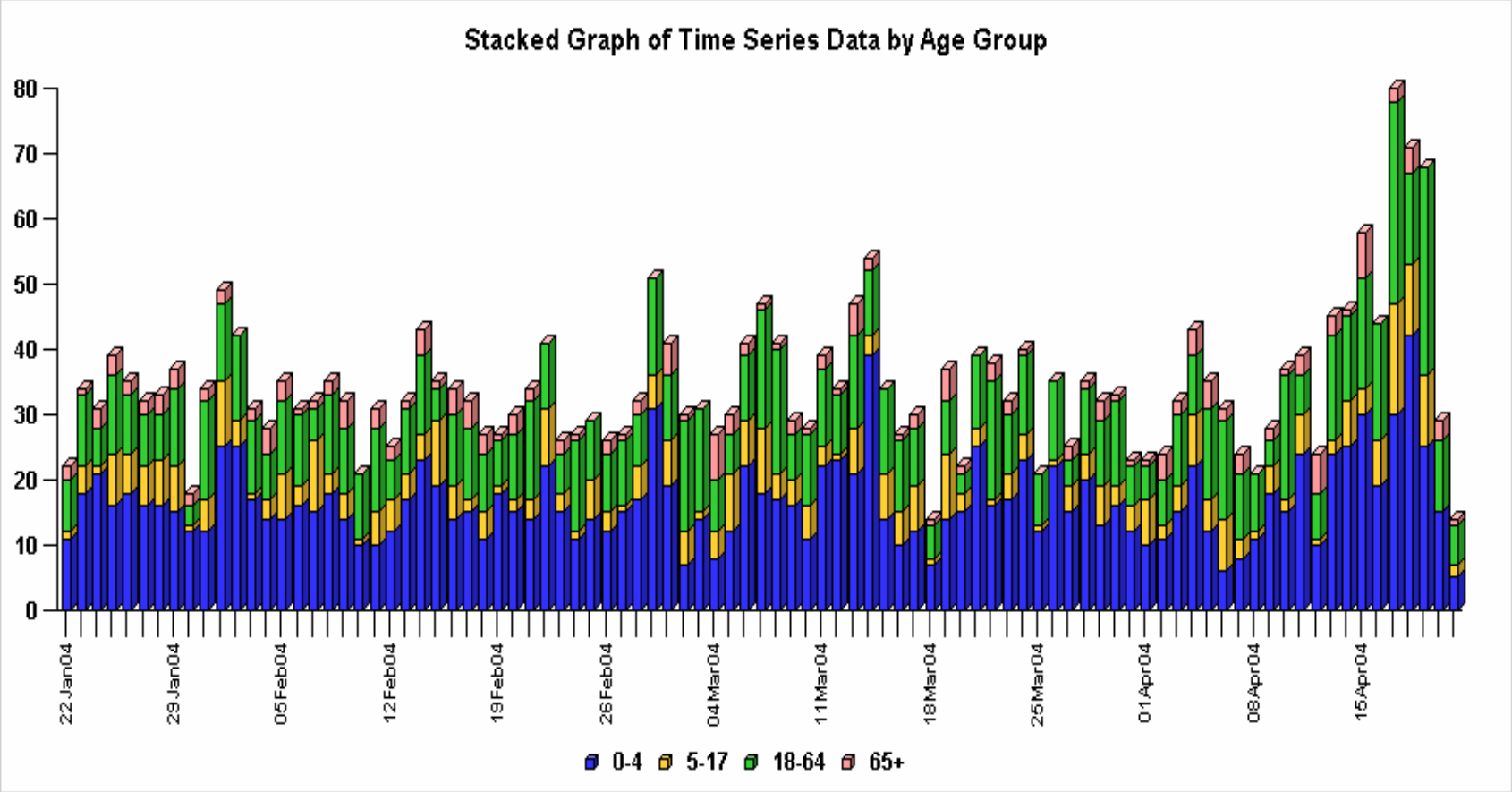
Sample Query Function



Popup Options

[Time Series | Data Details | Map View]

ESSENCE IV - Demo Time Series - Microsoft Internet Explorer provided by ITSD



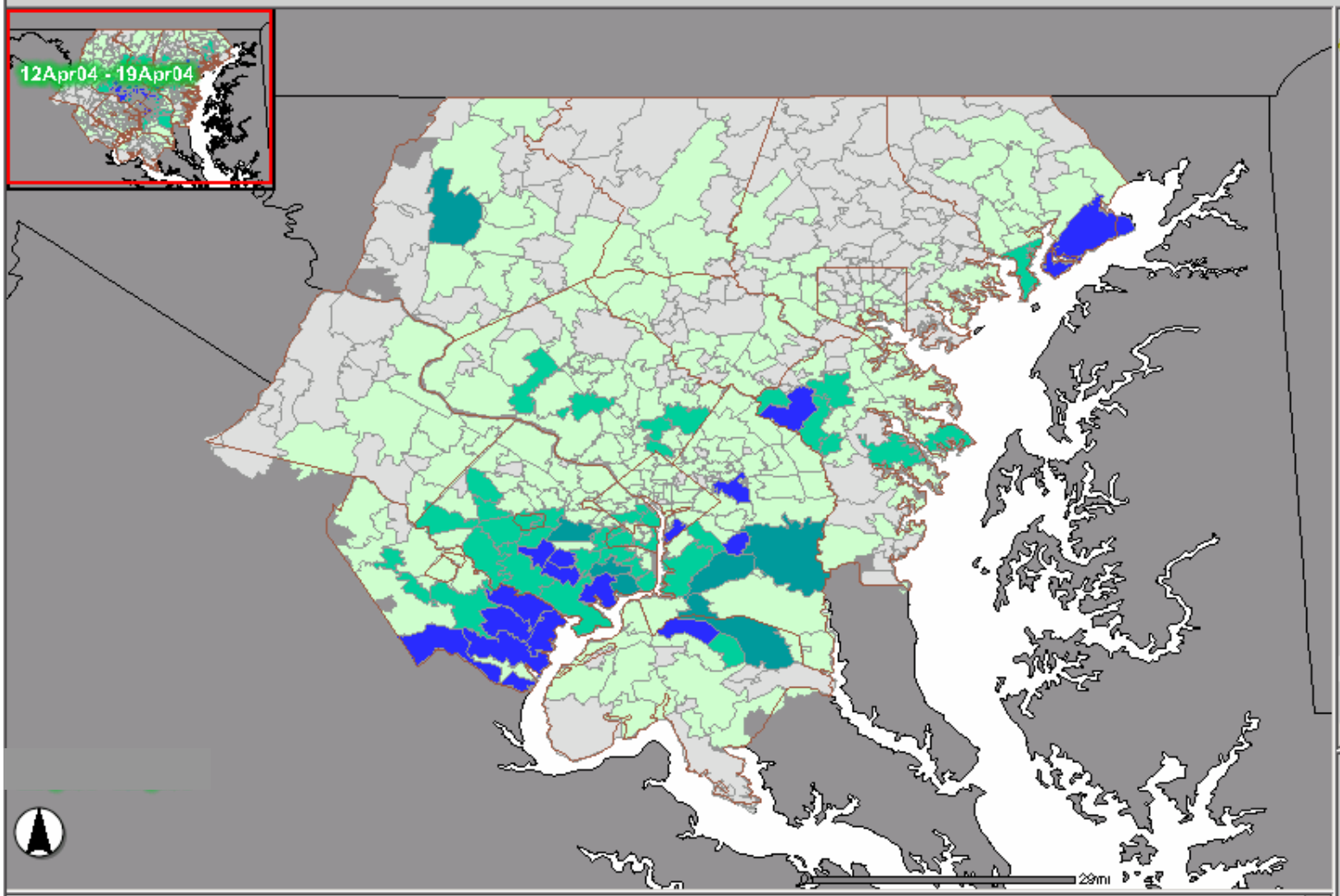
13Apr04	45	26	45	1.208	Data Details	Map View
12Apr04	24	20	24	0	Data Details	Map View
11Apr04	39	30	39	0.579	Data Details	Map View
10Apr04	37	33	37	0	Data Details	Map View

Internet

ESSENCE IV - NCR - Microsoft Internet Explorer provided by ITSD

ESSENCE IV - NCR : Data Sources Map

Toggles: Navigation: Date Range: Other: Help: Current Tool: Zoom In



LAYERS

- All Layers
  - Base Layers
  - School Absences
  - OTC Sales
    - |Visible|Label|Active|Name|
    - otc1
    - otc1_rtd
    - otc2
    - otc2_rtd
    - otc3
    - otc3_rtd
  - ER By Hospital
    - |Visible|Label|Active|Name|
    - hosp
    - hosp%
    - hosp_rtd
  - ER By Zipcode
    - |Visible|Label|Active|Name|
    - er
    - er_rtd
  - Civilian Office Visits
  - Elderly Office Visits
  - Military Outpatient Visits
    - |Visible|Label|Active|Name|
    - mil01
    - mil01_rtd

Legend

- Counties
- mil01
  - N/A
  - 0-15
  - 16-39
  - 40-63
  - 64+
- Zips
- States

Layer Name	mil01	**Data Column Name**	MIL_TOTAL
Data Source	Military Outpatient Visits	**Geography System**	Zipcode
Medical Grouping System	ESSENCE Syndrome	**ESSENCE Syndrome**	All
Age Group	All	**Sex**	All
Detector	JHUAPL-Regression/EWMA	**Start Date**	

Map: -75.67 , 38.71 -- Image: 813 , 360 -- ScaleFactor: 0.003314917127071883 Internet

Encounter Listings

History of ESSENCE | Syndrome Definitions | Detector Algorithms | Data Dictionary | Tutorial

HOME | Alert List | Query | Map Portal | Message Board

ESSENCE IV - Demo Data Details

[+] Description
[+] Configuration Options
[−] Available Links

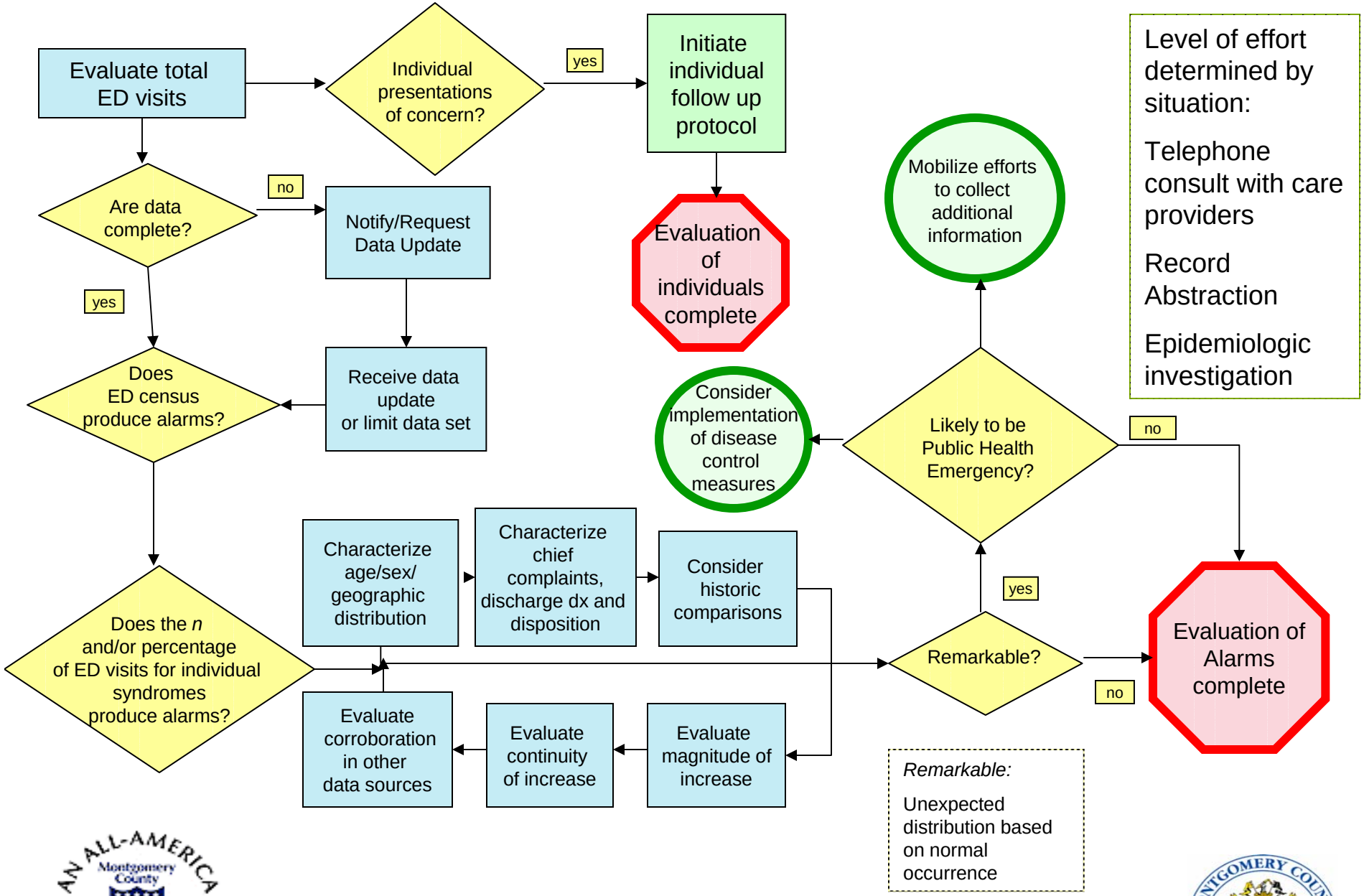
[Time Series | Data Details | Map View]

[Plain Text | Microsoft Excel]

Reset 3-Level Sorting

Date	Time	Hospital	HospitalName	HospitalZipCode	HospitalRegion	MedRecNo	Zipcode	Orig_Zipcode	Region	Age	S
17Apr04	10:52 PM	DC2ERCC	DC Hospital 2	20012	Washington	259540	20705	20705	PRINCE GEORGES	16	F
17Apr04	08:38 PM	MD2ERCC	MD Hospital 2	20904	Montgomery	831288	20781	20781	PRINCE GEORGES	39	M
17Apr04	none	DC3ERCC	DC Hospital 3	20019	Washington	713348	20719	20719	PRINCE GEORGES	7	U
17Apr04	none	DC3ERCC	DC Hospital 3	20019	Washington	287889	20233	20233	PRINCE GEORGES	8	U
17Apr04	03:59 PM	MD2ERCC	MD Hospital 2	20904	Montgomery	1094561	20705	20705	PRINCE GEORGES	0	M
17Apr04	09:40 PM	MD2ERCC	MD Hospital 2	20904	Montgomery	168626	20741	20741	PRINCE GEORGES	21	M
17Apr04	11:09 PM	MD2ERCC	MD Hospital 2	20904	Montgomery	606408	20781	20781	PRINCE GEORGES	0	F
17Apr04	01:03 PM	MD2ERCC	MD Hospital 2	20904	Montgomery	626068	20741	20741	PRINCE GEORGES	30	F
17Apr04	10:00 AM	MD2ERCC	MD Hospital 2	20904	Montgomery	1096662	20719	20719	PRINCE GEORGES	0	F

Internet



Courtesy of Kathy Hurt-Mullen MPH

Investigation of signals

- Descriptive statistics (patients in the signal)
- Check complementary systems, compare to previous signals
- Acquire interim data (12-hour log)
- Calls to assess current situation
- Chart reviews
- Patient follow-up (from signal)
- Augment lab testing (new patients)
- On site teams (interviews, lab testing)

Courtesy of Don Weiss MD, MPH

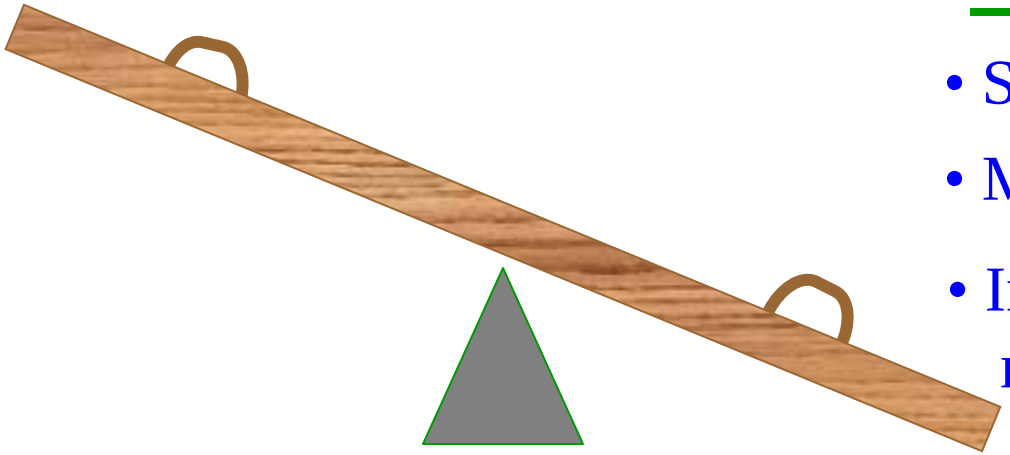
NYC Health

Constraints

Variables that make it difficult to come up with a common protocol

1) Size of the Health Department
   - Resources available at the time to dedicate to the alert
2) Amount of data available to indicate intervention
3) Threat level or degree of risk

Determining the threshold

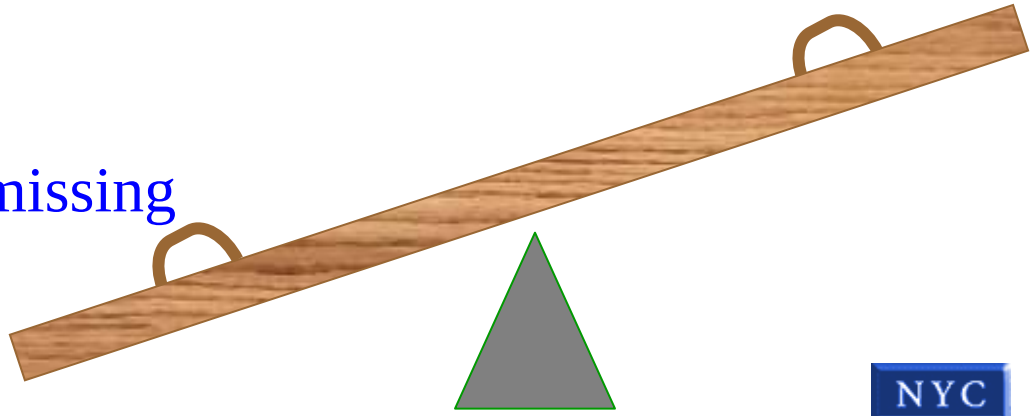


Favor specificity

- Set p value low (0.001)
- Minimize false alarms
- Increase probability of missing a real event

Favor sensitivity

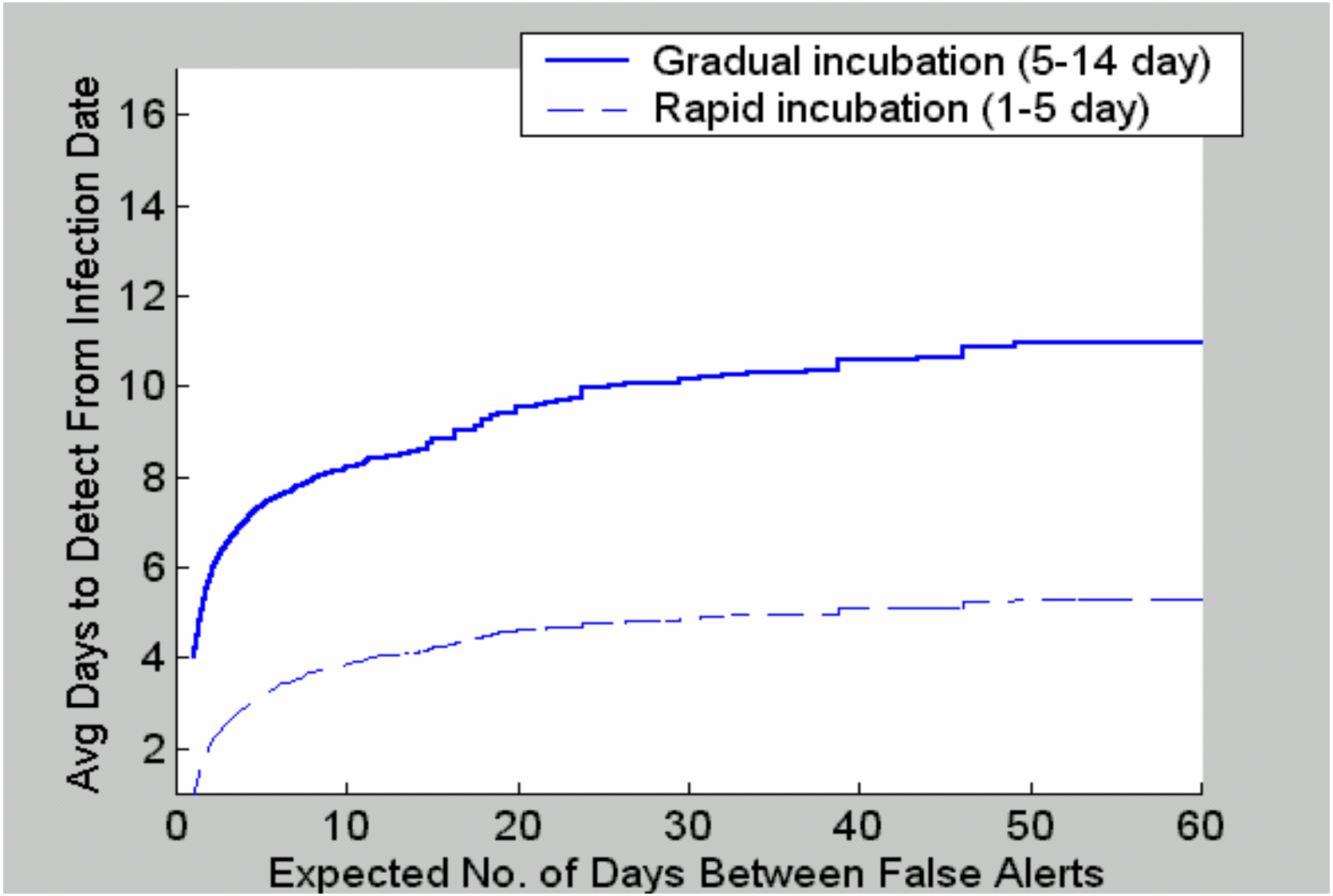
- Set p value high (0.05)
- Decrease probability of missing a real event
- Many false alarms



Courtesy of Don Weiss MD, MPH

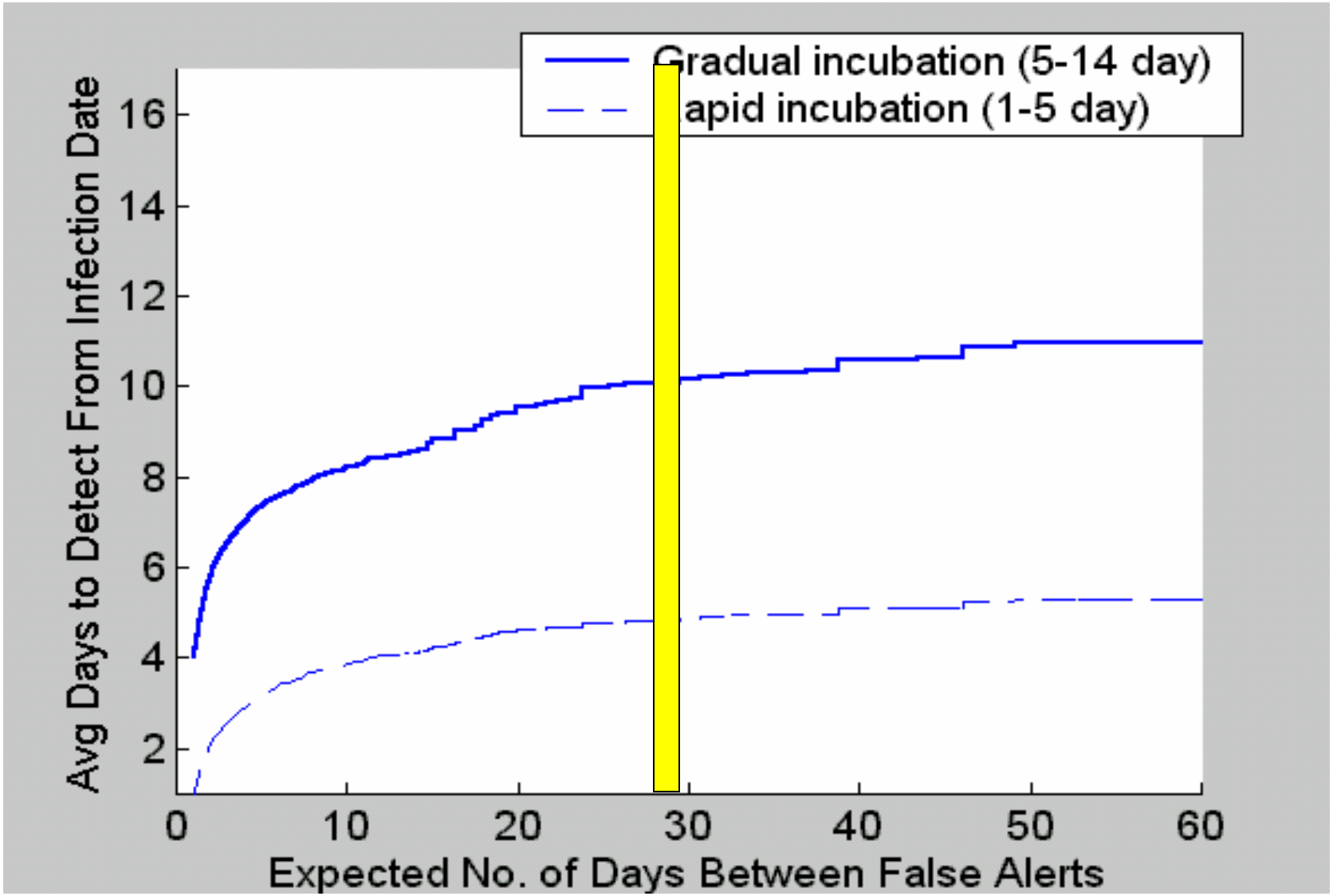
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



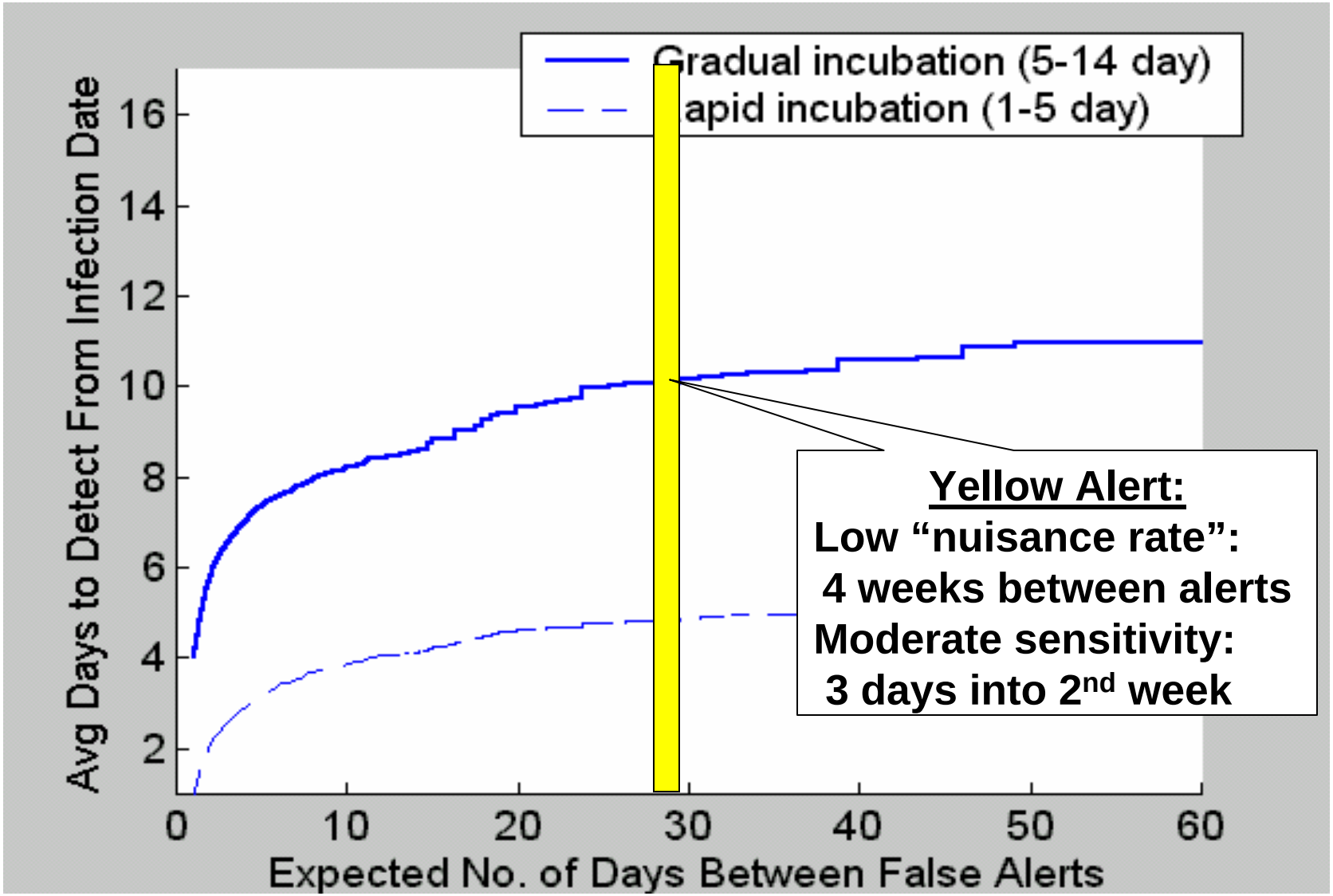
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



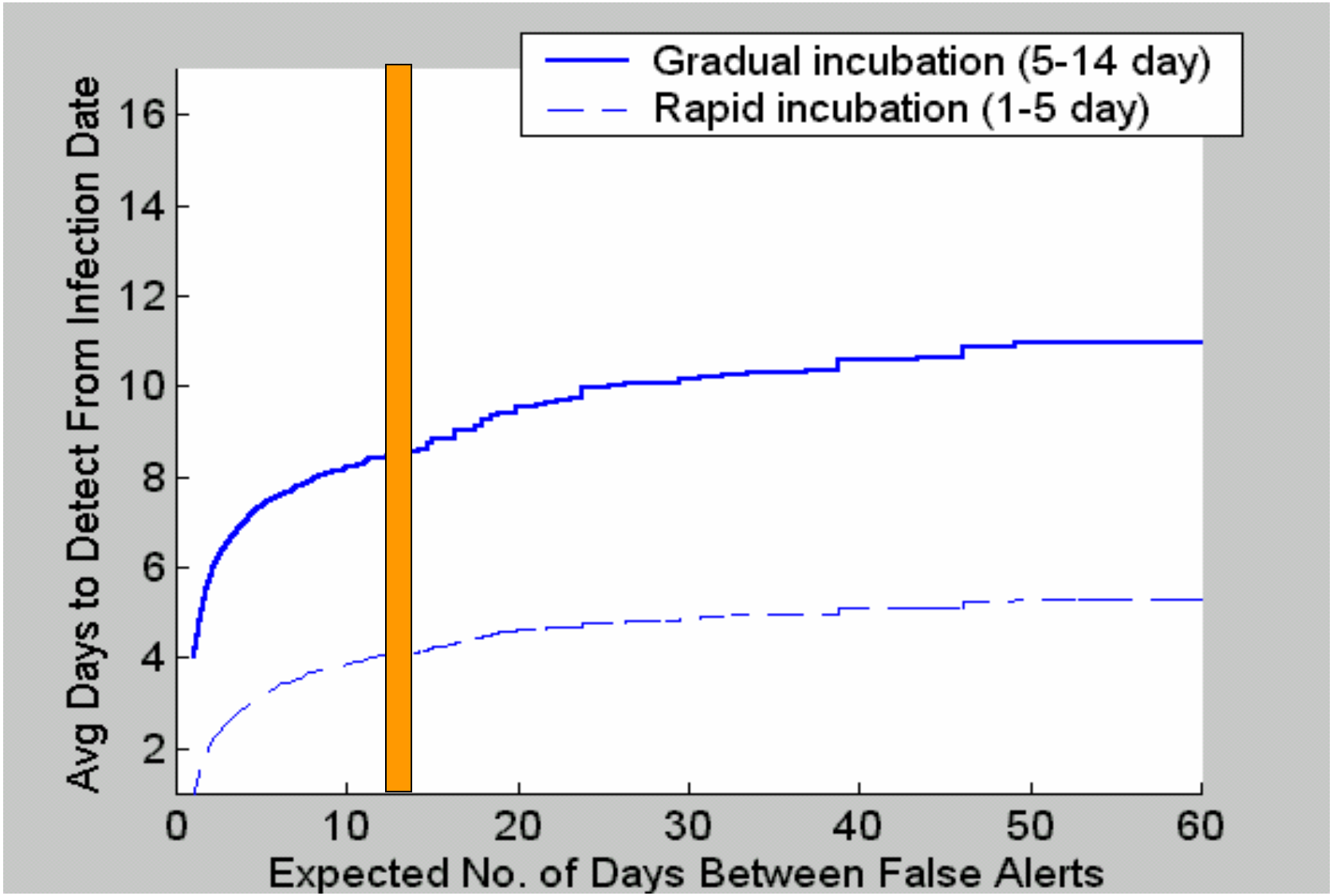
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



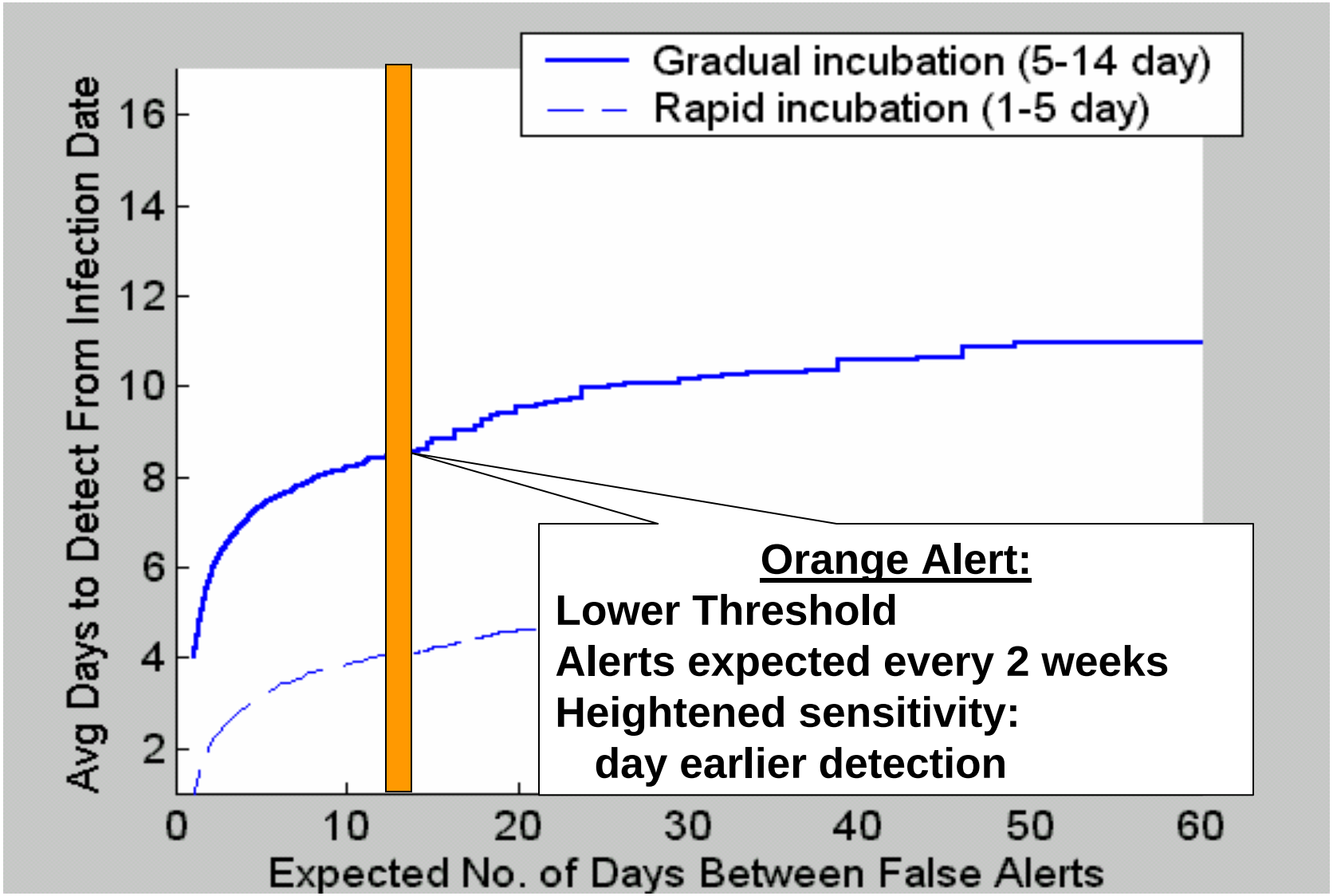
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



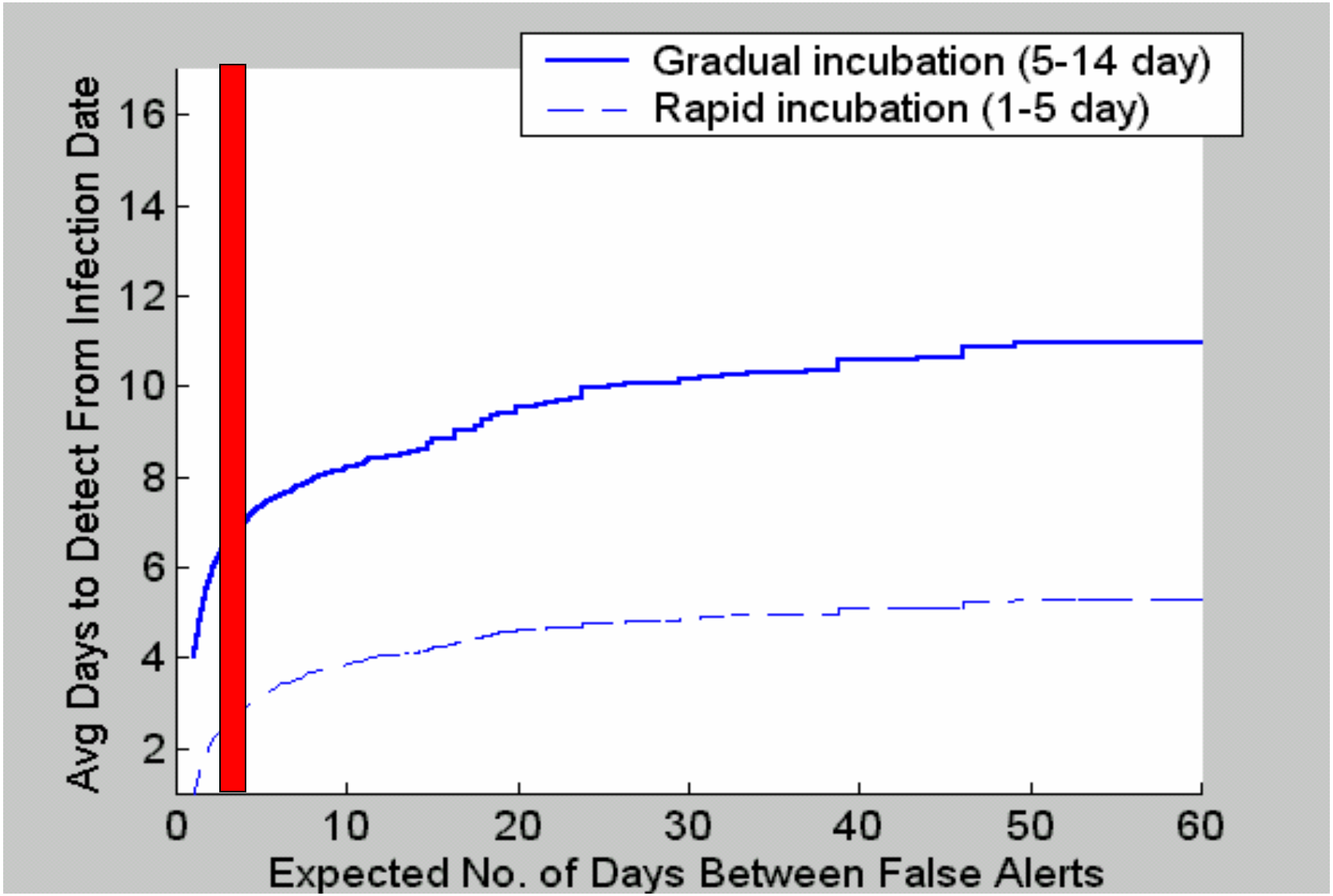
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



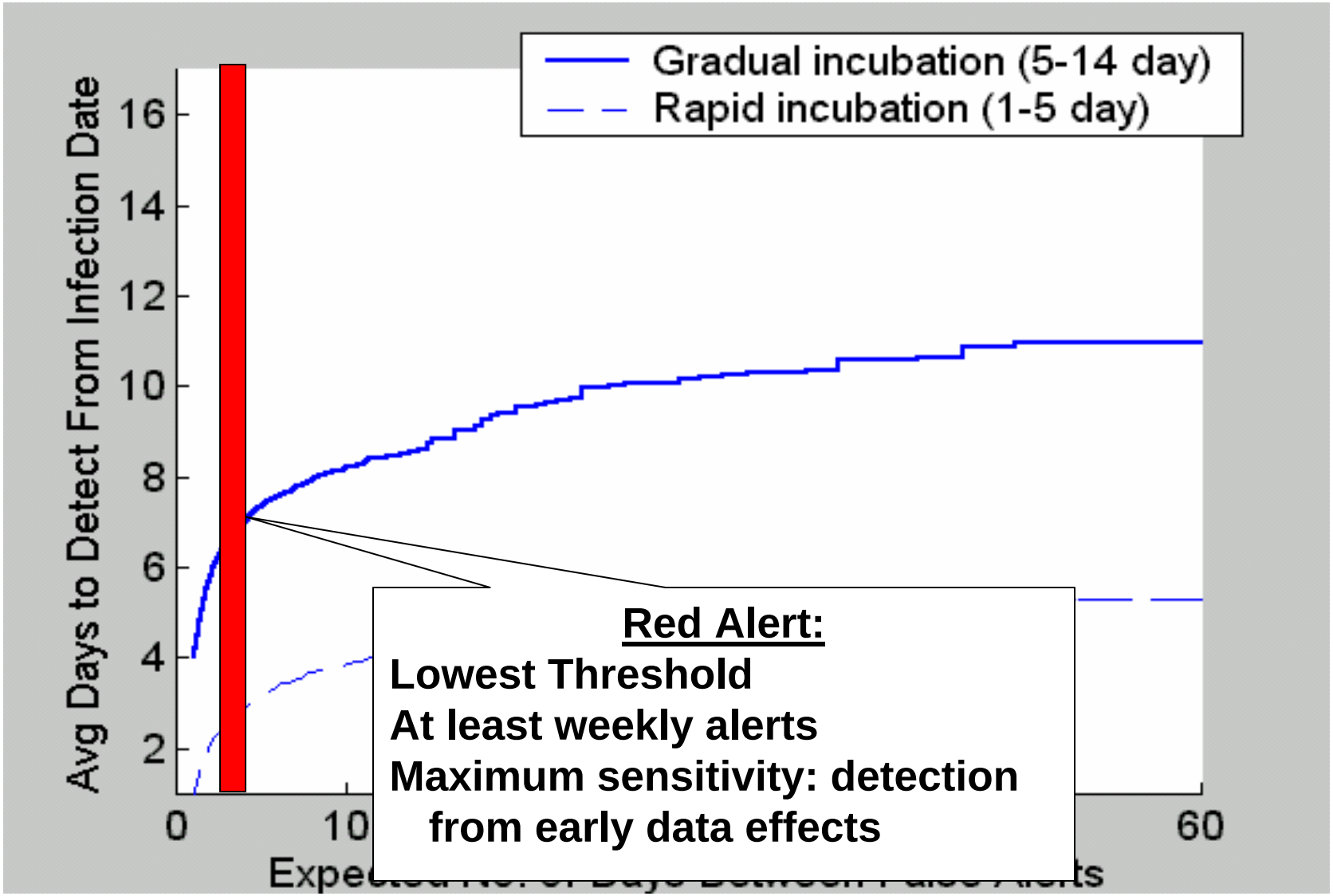
Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



Role of Alerting Algorithms In Outbreak Investigation

Thresholds and Awareness Levels



Additional Requirements to Assist the Epidemiologist

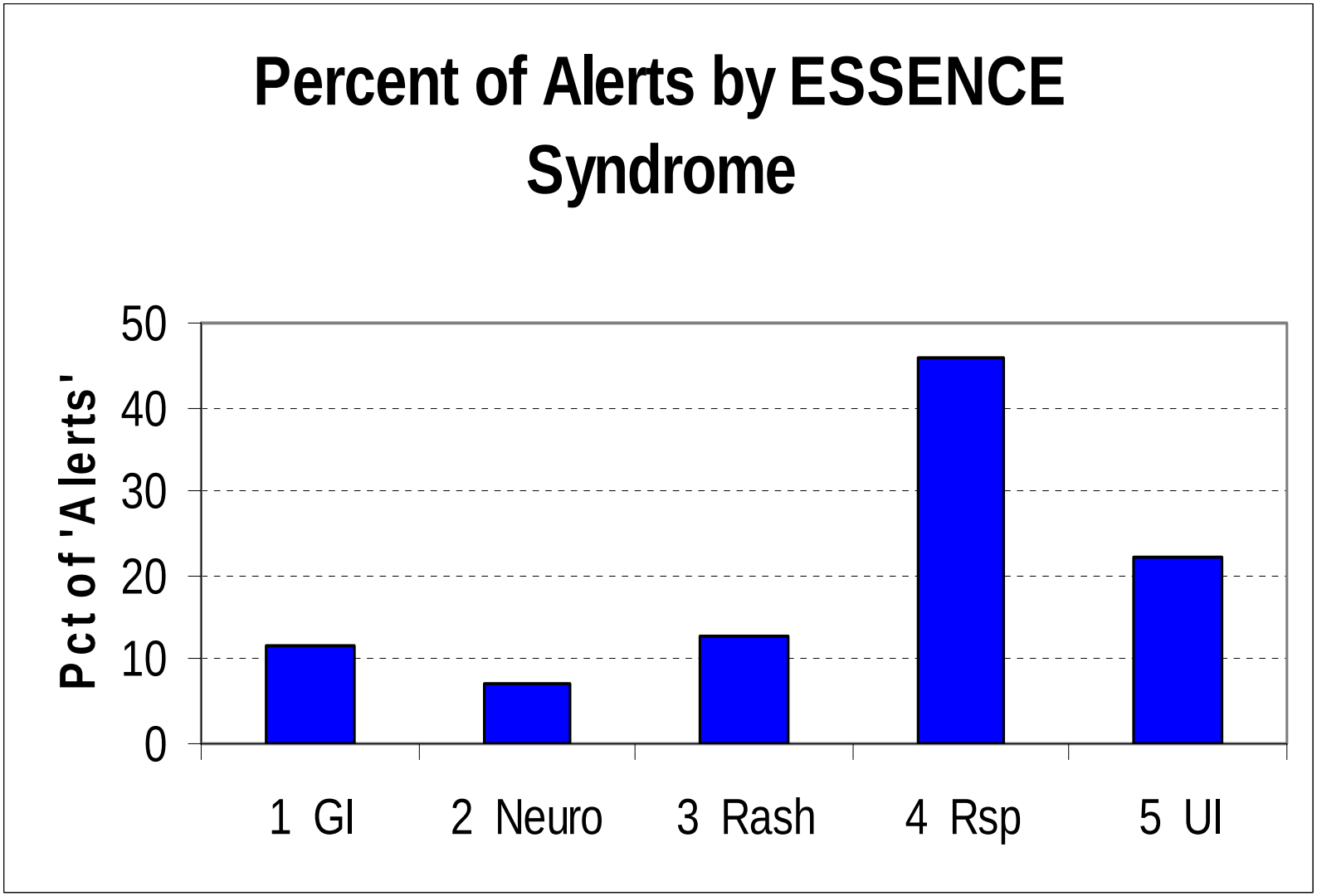
Data Representation Functions:

1) Sorting
2) Frequency distributions
3) Cross tabulations of data
4) Additional graphing functions

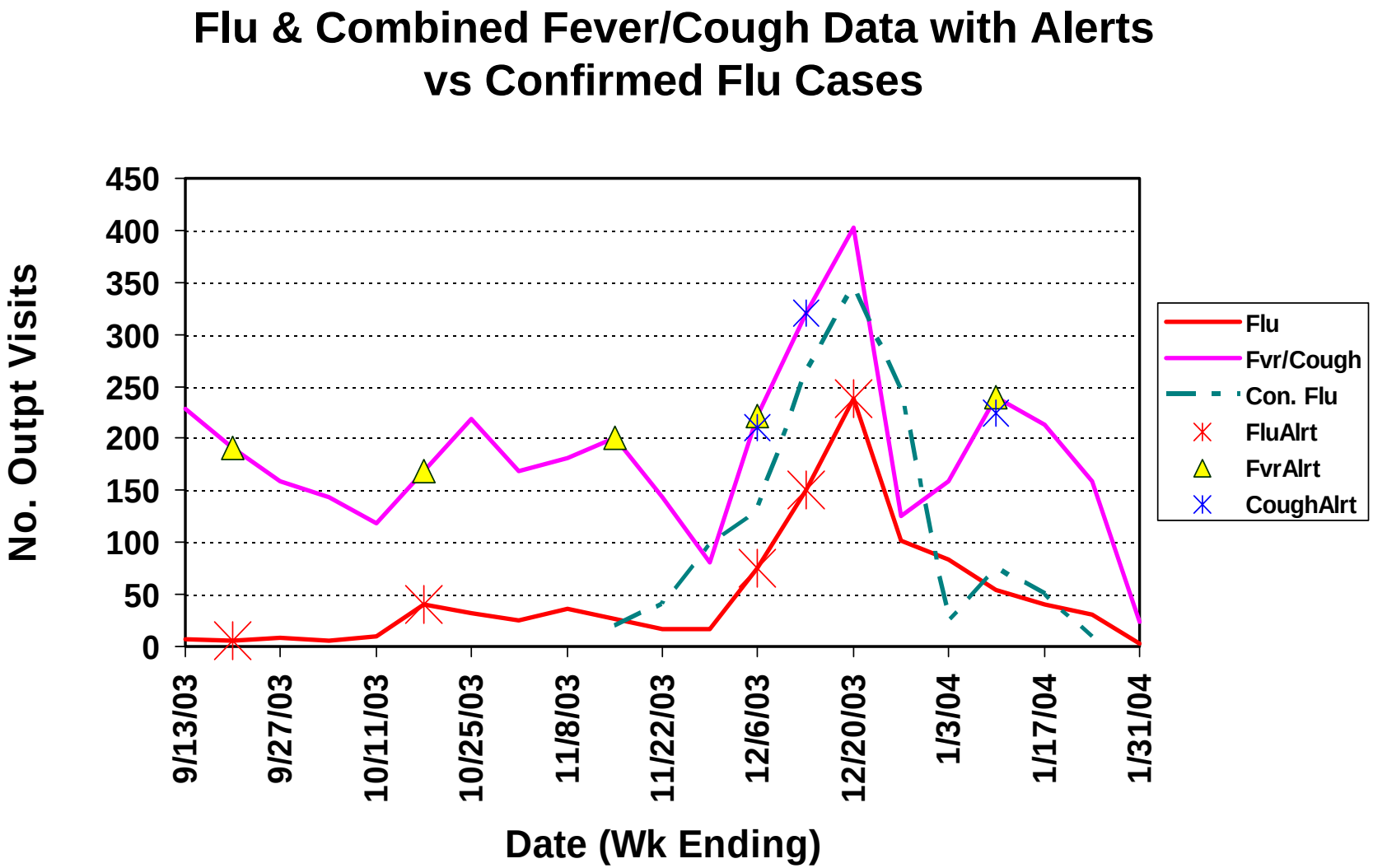
Data Transfer Functions:

1) Excel , SAS, & other formats for analysis
2) ebXML / HL-7 export / import, to / from other surveillance systems
3) Etc.

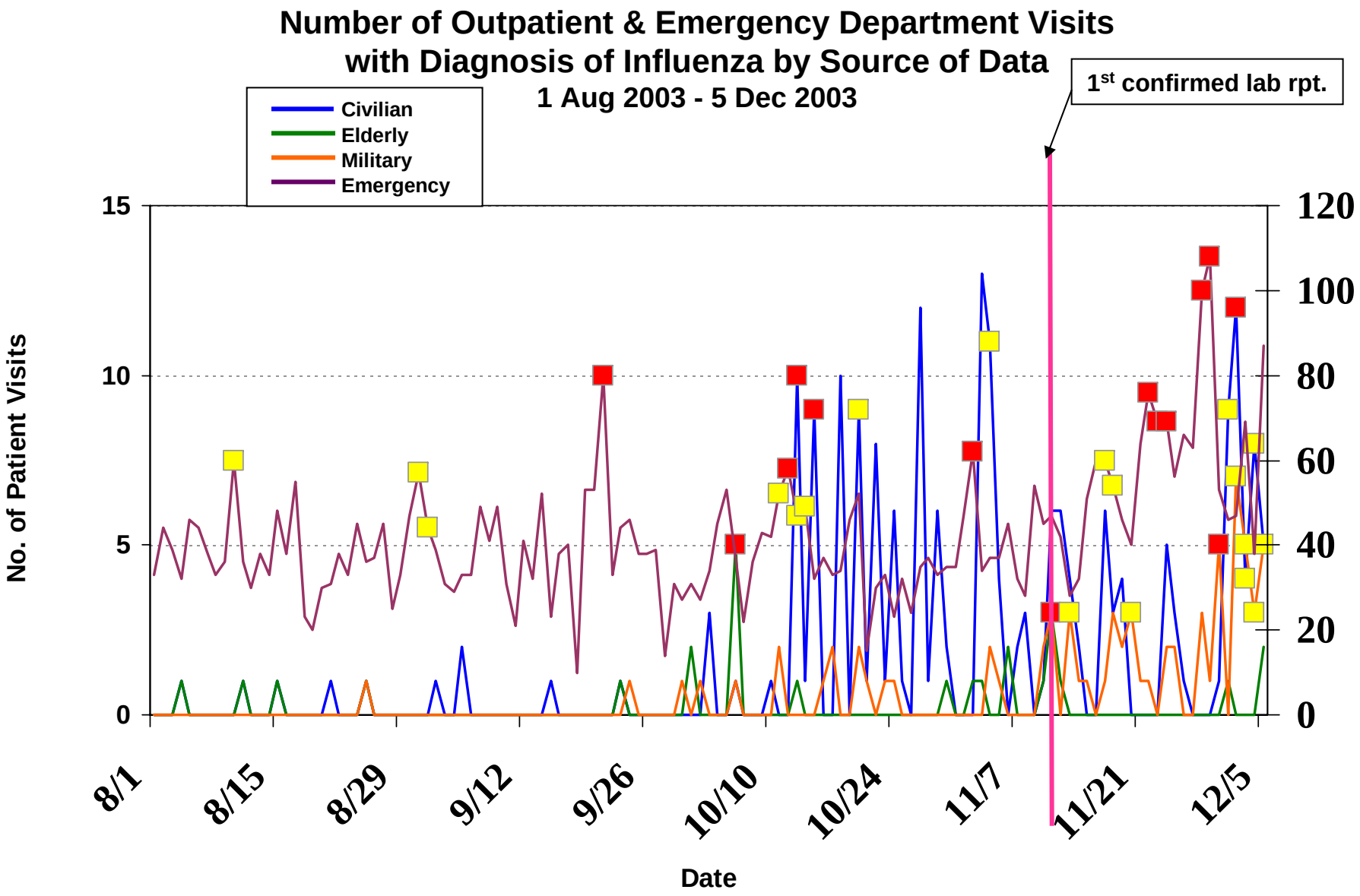
Frequency Distributions



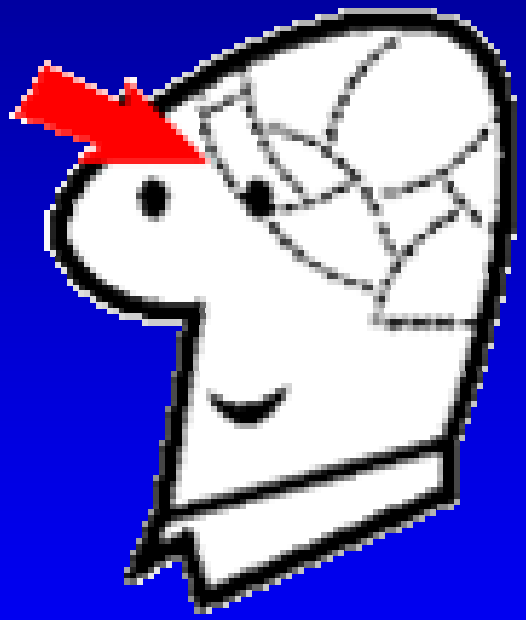
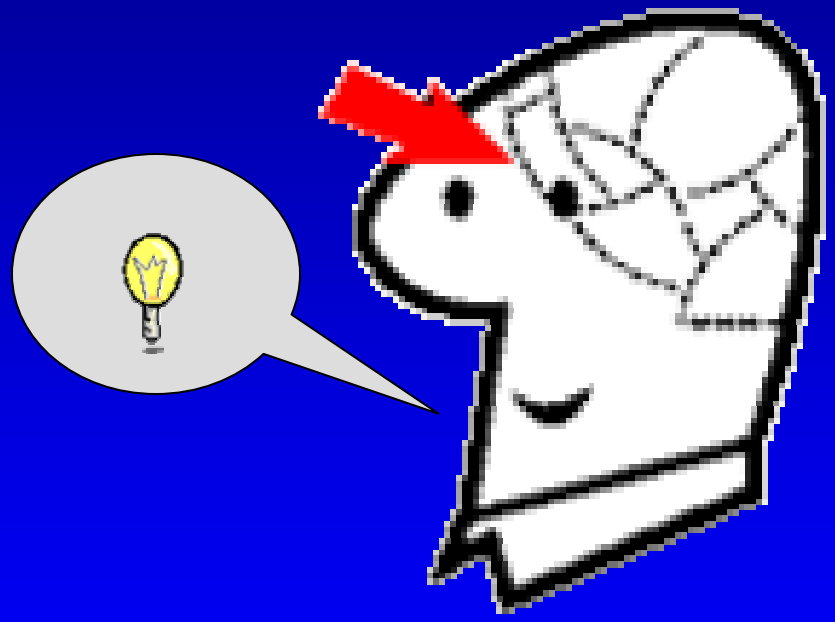
Graphed Multiple Data Streams



Another Example of Multiple Data Graphing



STAKEHOLDER THOUGHTS



AND INPUTS