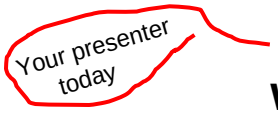
Rule-Based Anomaly Pattern Detection for Detecting Disease Outbreaks



Andrew Moore
Weng-Keen Wong
Mike Wagner
Greg Cooper

In collaboration with many others, including Wendy Chapman, Bill Hogan, Bob Olszewski, Jeff Schneider , Rich Tsui, Weng-Keen Wong

University of Pittsburgh and Carnegie Mellon University
www.cs.cmu.edu/~awm

The Auton Lab

Faculty:	Andrew Moore, Jeff Schneider, Artur Dubrawski
Postdoctoral Fellows:	Brigham Anderson, Alexander Gray, Paul Komarek, Dan Pelleg
Graduate Students:	Brent Bryan, Kaustav Das, Khalid El-Arini, Anna Goldenberg, Jeremy Kubica, Ting Liu, Daniel Neill, Sajid Siddiqi, Purna Sarkar, Ajit Singh, Weng-Keen Wong
Head of Software Development:	Jeanie Komarek
Programmers:	Patrick Choi, Adam Goode, Pat Gunn, Joey Liang, John Ostlund, Robin Sabhnani, Rahul Sankathar
Executive Assistant:	Kristen Schrauder
Head of Sys. Admin:	Jacob Joseph
Undergraduate and Masters Interns:	Kenny Daniel, Sandy Hsu, Dongryeol Lee, Jennifer Lee, Avilay Parekh, Chris Rotella, Jonathan Terleski
Recent Alumni:	Drew Bagnell (RI faculty), Scott Davies (Google), David Cohn (Google), Geoff Gordon (CMU), Paul Hsiung (USC), Marina Meila (U. Washington), Remi Munos (Ecole Polytechnique), Malcolm Strens (Qinetiq)

Current Sponsors

- National Science Foundation (NSF)
- NASA
- Defense Advanced Research Projects Agency (DARPA)
- Other Government Sponsor
- Department of Homeland Security (DHS)
- Homeland Security Advanced Research Projects Agency (HSARPA)
- United States Department of Agriculture (USDA)

Federal

- Health Canada
- State of Pennsylvania

State

- A Fortune-50 Pharmaceutical company
- Caterpillar Corporation
- A Fortune-50 Petrochemical company

Large industrial

- Psychogenics Corporation
- Transform Pharmaceuticals

Small industrial

Collaborators...


Ting Liu


Paul Hsiung


Jeff Schneider


Alex Gray


Brigham Anderson


Kan Deng


Anna Goldenberg


Daniel Neill


Scott Davies

Ajit Singh


Jeremy Kubica


Paul Komarek


Weng-Keen Wong


Dan Pelleg

Drew Bagnell

Sajid Siddiqi

Andy Connolly

Greg Cooper

Mike Wagner

Andrew Lawson

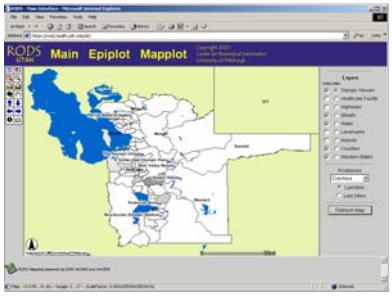
Rich Tsui

1996	1997	1998	1999	2000	2001	2002	2003
M&M Mars Line control Adrenaline (NOX mini-malization)	Kodak (Image sta-bilization) Digital Equipment (pregnan-cy monit-oring)	M&M Mars (manufa c-turing) NASA/ NSF (Astrop hysics mining) 3M: Textile tension control	Caterpillar (Spare parts) US Army (biotoxin detection) M&M Mars: Scheduling with uncer-tainty 3M (Adhesive design)	DigitalMC (Music tastes) Caterpillar (emissions) SmartMoney (anomalies) Unilever (Brand Management) Phillips Petroleum (work-force optimization) Cellomics (screened anomaly detection)	Biometrics company (health monitor) Boeing (intrusion) Masterfoods (new product development) Cellomics (pro-teomics screen) ABB (Circuit-breaker supply chain) SwissAir (Flight delays) 3M (secret) Washington Public Hospital System (ER delays) Unilever (targeted marketing)	NASA (National Virtual Observatory) NSF (astrostatistics software) DARPA (national disease monitor) Masterfoods (bio-chemistry) Pfizer (High-throughput screen) Caterpillar Inc. (Self-optimizing Engines) Beverage Company (Ingredients/Manufacturing/Marketing/ Sales Bayes Net) Transform Pharma (massive autonomous experiment design) Census Bureau (privacy protection) Psychogenics Inc: Effects of psychotropic drugs on rats	NSF (astrostatistics software) Masterfoods (bio-chemistry) State of PA (National Disease Monitor [with Mike Wagner of U. Pitt]) State of PA (Anti Cancer [collaboration with CMU Biology] DARPA (detecting patterns in links) Other Government Departments (identifying dangerous people, potential collaborators, and aliases) Other Government Departments (detecting a class of clusters) Other Pharma Research Co. Life Science specific data mining United States Department of Agriculture: Early warning system for food terrorism NSF: Biosurveillance Algorithms

Auton/SPR Deployments

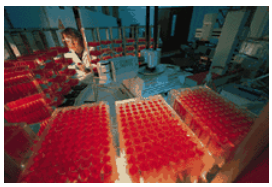
Our 5 biggest applications in 2004

Biomedical Security (with Mike Wagner, University of Pittsburgh)



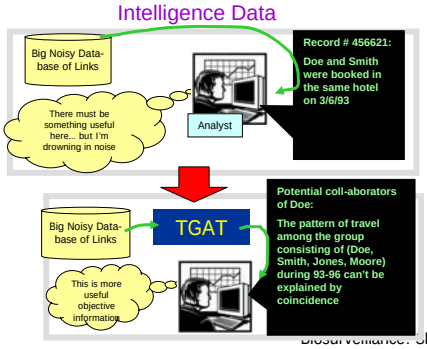
Autonomous self-tweaking engines

Drug Screening



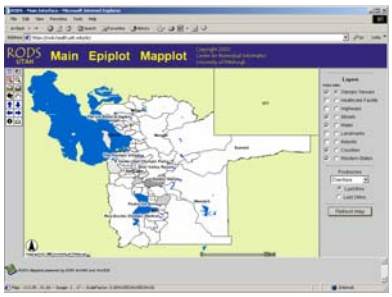
Big Astrophysics Automated Science

Intelligence Data



Our 5 biggest applications in 2002

Biomedical Security (with Mike Wagner, University of Pittsburgh)

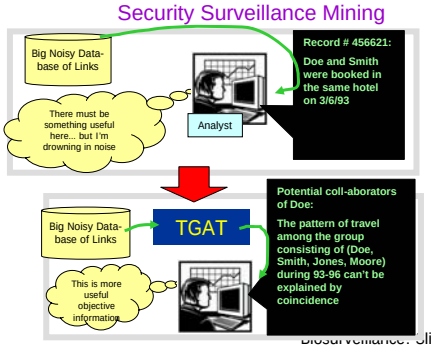


Autonomous self-tweaking engines

Drug Screening

Big Astrophysics Automated Science

Security Surveillance Mining

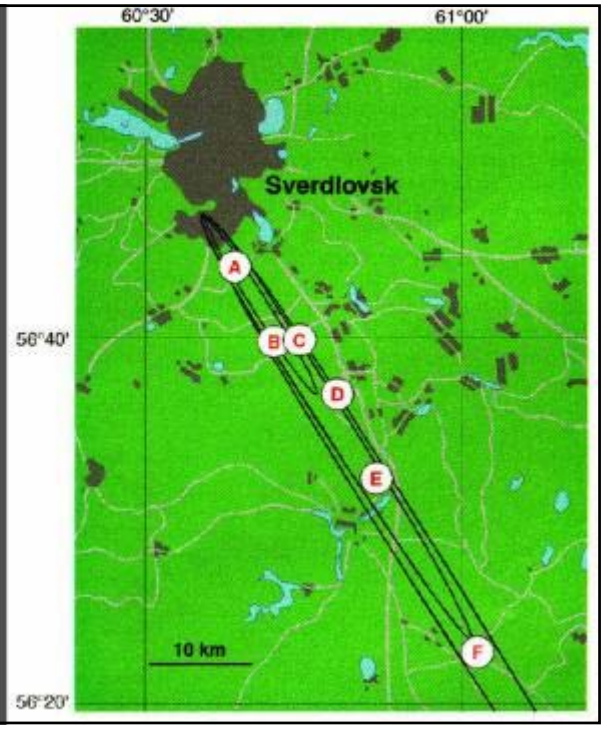


..Early Thursday Morning. Russia. April 1979...

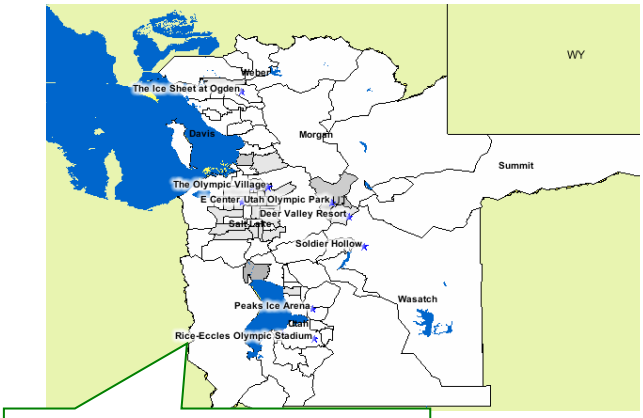
Sverdlovsk: Aerial View

- During April and May 1979, there were 77 Confirmed cases of inhalational anthrax

Sverdlovsk Region: Epi-map

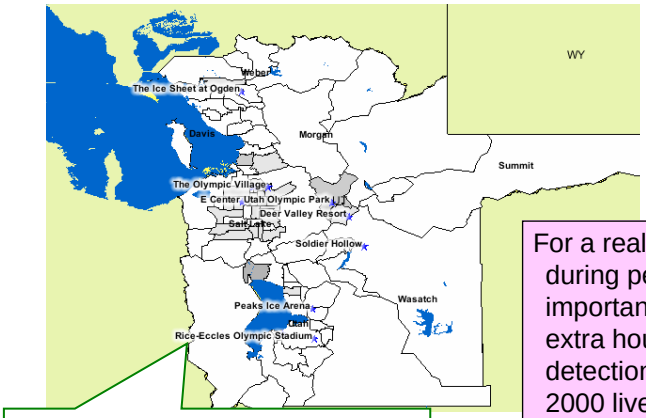


What if this happened in 2002?



Salt-lake city, Feb 2002

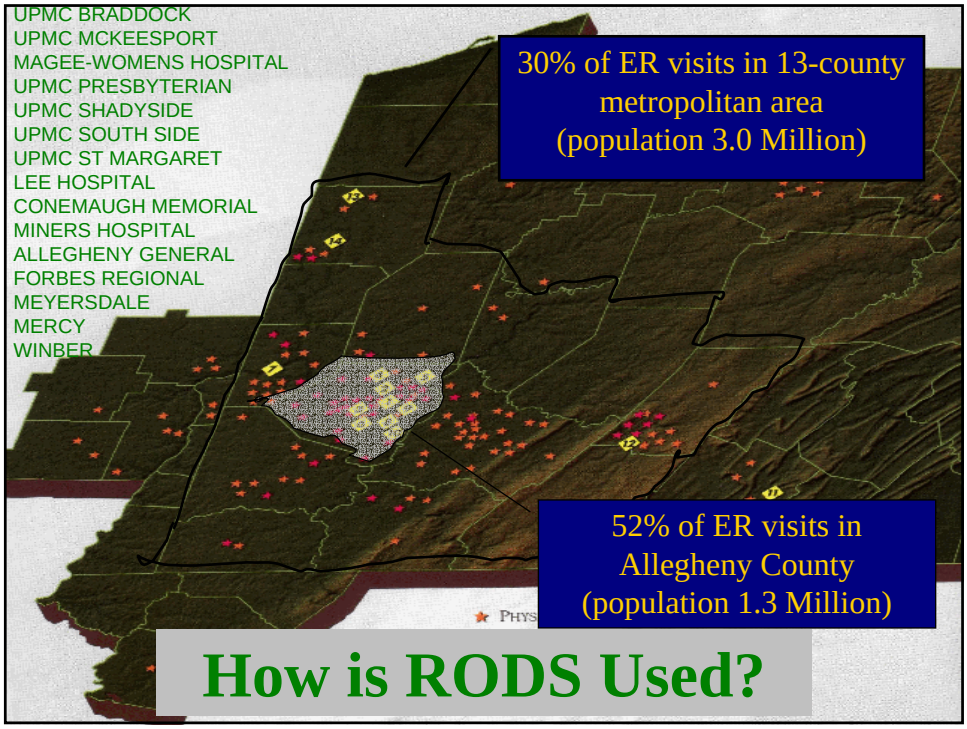
What if this happened in 2002?



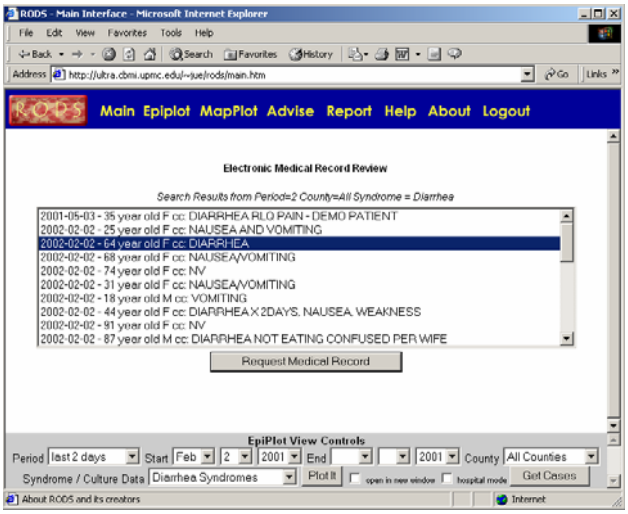
Salt-lake city, Feb 2002

For a realistic attack, during peak period of importance, each extra hour of early detection could save 2000 lives.

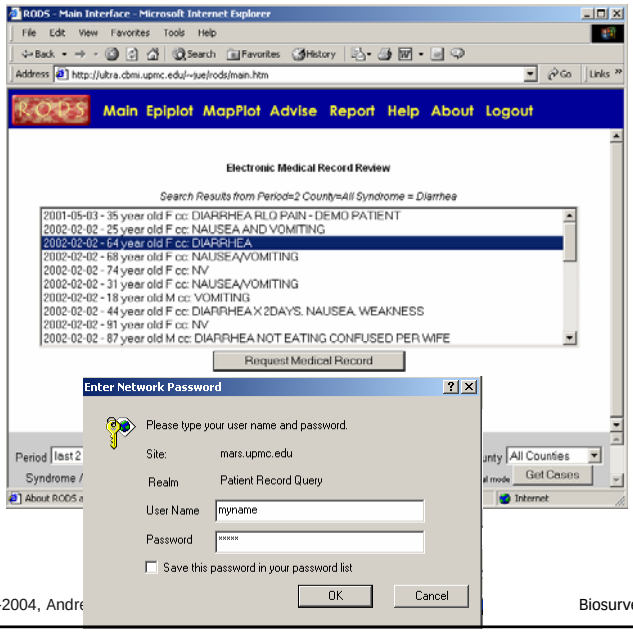
[Tsui et al., 2000]



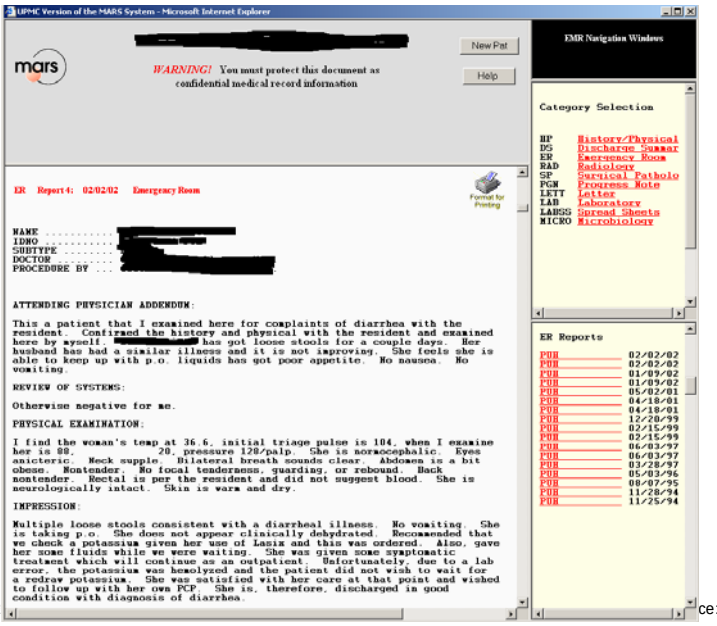
When an alarm sounds...



When an alarm sounds...



How is RODS Used?



The task

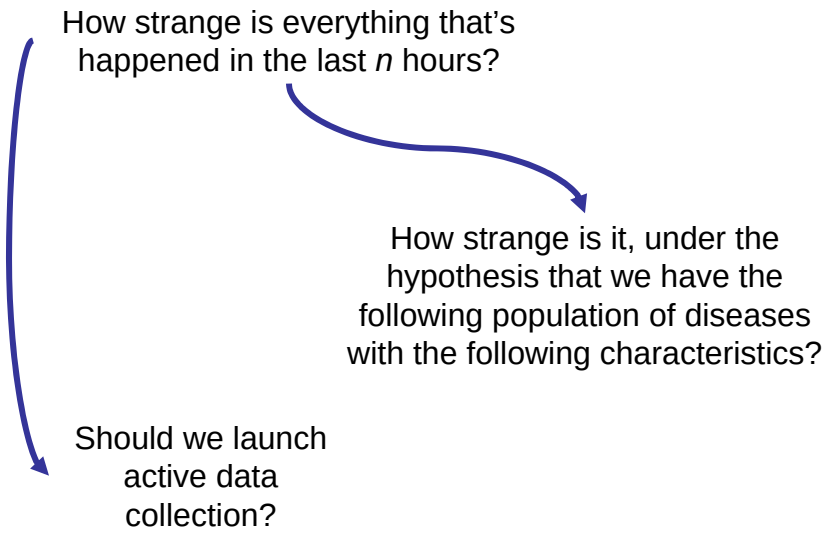
How strange is everything that's happened in the last *n* hours?

The task

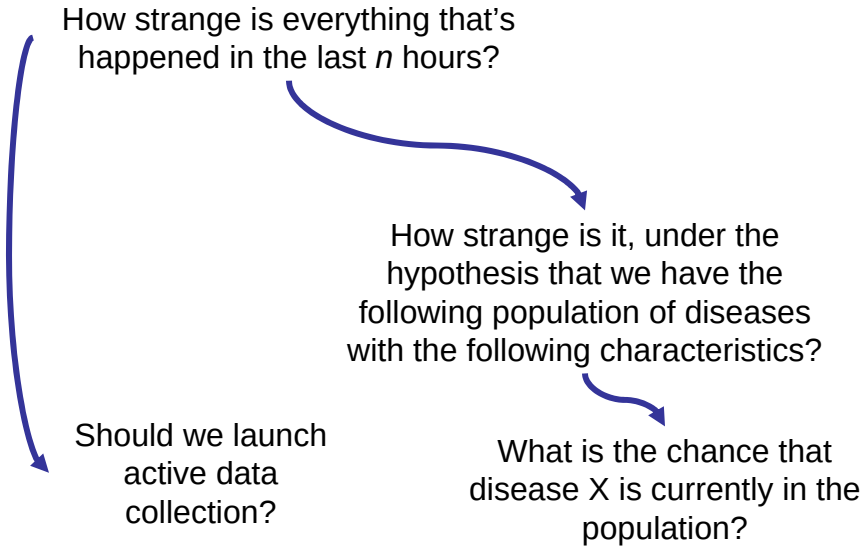
How strange is everything that's happened in the last *n* hours?

Should we launch active data collection?

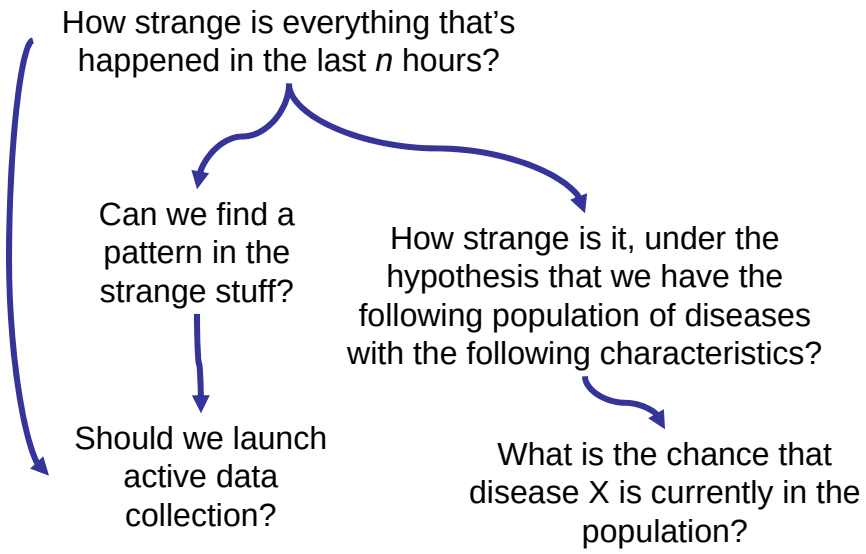
The task



The task



The task



WSARE v2.0

- What's Strange About Recent Events?
- Designed to be easily applicable to any multivariate many-attribute date/time-indexed biosurveillance-relevant data stream.

WSARE v2.0

- Inputs:
  - 1. Date/time-indexed biosurveillance-relevant data stream
  - 2. Time Window Length
  - 3. Which attributes to use?

WSARE v2.0

- Inputs:
  - 1. Date/time-indexed biosurveillance-relevant data stream
  - 2. Time Window Length
  - 3. Which attributes to use?

Example

"last 24 hours"

"ignore key and weather"

Primary Key	Date	Time	Hospital	ICD9	Prodrome	Gender	Age	Home			Work			Recent Flu Levels	Recent Weather	(Many more...)
								Large Scale	Medium Scale	Fine Scale	Large Scale	Medium Scale	Fine Scale			
h6r32	6/2/2	14:12	Down-town	781	Fever	M	20s	NE	15217	A5	NW	15213	B8	2%	70R	...
t3q15	6/2/2	14:15	River-side	717	Respirat ory	M	60s	NE	15222	J3	NE	15222	J3	2%	70R	...
t5hh5	6/2/2	14:15	Smith-field	622	Respirat ory	F	80s	SE	15210	K9	SE	15210	K9	2%	70R	...
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

WSARE v2.0

- Inputs:
  - 1. Date/time-indexed biosurveillance-relevant data stream
  - 2. Time Window Length
  - 3. Which attributes to use?
- Outputs:
  - 1. Here are the records that most surprise me
  - 2. Here's why
  - 3. And here's how seriously you should take it

Primary Key	Date	Time	Hospital	ICD9	Prodrome	Gender	Age	Home			Work			Recent Flu Levels	Recent Weather	(Many more...)
								Large Scale	Medium Scale	Fine Scale	Large Scale	Medium Scale	Fine Scale			
h6r32	6/2/2	14:12	Down-town	781	Fever	M	20s	NE	15217	A5	NW	15213	B8	2%	70R	...
t3q15	6/2/2	14:15	River-side	717	Respiratory	M	60s	NE	15222	J3	NE	15222	J3	2%	70R	...
t5hh5	6/2/2	14:15	Smith-field	622	Respiratory	F	80s	SE	15210	K9	SE	15210	K9	2%	70R	...
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

Simple WSARE

- Given 500 day's worth of ER cases at 15 hospitals...

Date	Cases
Thu 5/22/2000	C1, C2, C3, C4 ...
Fri 5/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Sat 12/9/2000	C1, C2, C3, C4 ...
Sun 12/10/2000	C1, C2, C3, C4 ...
:	:
Sat 12/16/2000	C1, C2, C3, C4 ...
:	:
Sat 12/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Fri 9/14/2001	C1, C2, C3, C4 ...

Simple WSARE

- Given 500 day's worth of ER cases at 15 hospitals...
- For each day...
  - Take today's cases

Date	Cases
Thu 5/22/2000	C1, C2, C3, C4 ...
Fri 5/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Sat 12/9/2000	C1, C2, C3, C4 ...
Sun 12/10/2000	C1, C2, C3, C4 ...
:	:
Sat 12/16/2000	C1, C2, C3, C4 ...
:	:
Sat 12/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Fri 9/14/2001	C1, C2, C3, C4 ...

Simple WSARE

- Given 500 day's worth of ER cases at 15 hospitals...
- For each day...
  - Take today's cases
  - The cases one week ago
  - The cases two weeks ago

Date	Cases
Thu 5/22/2000	C1, C2, C3, C4 ...
Fri 5/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Sat 12/9/2000	C1, C2, C3, C4 ...
Sun 12/10/2000	C1, C2, C3, C4 ...
:	:
Sat 12/16/2000	C1, C2, C3, C4 ...
:	:
Sat 12/23/2000	C1, C2, C3, C4 ...
:	:
:	:
Fri 9/14/2001	C1, C2, C3, C4 ...

Simple WSARE

- Given 500 day's wor
  of ER cases at 15
  hospitals...
- For each day...
  - Take today's cases
  - The cases one week ago
  - The cases two weeks ago
- Ask: "What's different
  about today?"

DATE_AD	ICD9	PRODRON	GENDER	place2	…	…
					…	…
12/9/00	786.05	3	F	s-e	…	…
12/9/00	789	1	F	s-e	…	…
12/9/00	789	1	M	n-w	…	…
12/9/00	786.05	3	M	s-e	…	…
:	:	:	:	:	…	…
12/16/00	787.02	2	M	n-e	…	…
12/16/00	782.1	4	F	s-w	…	…
12/16/00	789	1	M	s-e	…	…
12/16/00	786.09	3	M	n-w	…	…
12/23/00	789.09	1	M	s-w	…	…
12/23/00	789.09	1	F	s-w	…	…
12/23/00	782.1	4	M	n-w	…	…
:	:	:	:	:	…	…
12/23/00	786.09	3	M	s-e	…	…
12/23/00	786.09	3	M	s-e	…	…
12/23/00	780.9	2	F	n-w	…	…
12/23/00	V40.9	7	M	s-w	…	…

Simple WSARE

- Given 500 day's wor
  of ER cases at 15
  hospitals...
- For each day...

DATE_AD	ICD9	PRODRON	GENDER	place2	…	…
					…	…
12/9/00	786.05	3	F	s-e	…	…
12/9/00	789	1	F	s-e	…	…
12/9/00	789	1	M	n-w	…	…
12/9/00	786.05	3	M	s-e	…	…
:	:	:	:	:	…	…
12/16/00	787.02	2	M	n-e	…	…
12/16/00	782.1	4	F	s-w	…	…
12/16/00	789	1	M	s-e	…	…
12/16/00	786.09	3	M	n-w	…	…
12/23/00	789.09	1	M	s-w	…	…
		1	F	s-w	…	…
				n-w	…	…
				:	…	…
				s-e	…	…
				s-e	…	…
				n-w	…	…
				s-w	…	…

Fields we use:

Date, Time of Day, Prodrome, ICD9, ***Symptoms***, Age, Gender, Coarse Location, Fine Location, ***ICD9 Derived Features, Census Block Derived Features, Work Details, Colocation Details***

Example

```
Sat 12-23-2001 (daynum 36882, dayindex 239)

35.8% ( 48/134) of today's cases have 30 <= age < 40
17.0% ( 45/265) of   other cases have 30 <= age < 40
```


Example

```
Sat 12-23-2001 (daynum 36882, dayindex 239)
FISHER_PVALUE = 0.000051
35.8% ( 48/134) of today's cases have 30 <= age < 40
17.0% ( 45/265) of   other cases have 30 <= age < 40
```

Table 1: A sample 2x2 Contingency Table

	C_{today}	C_{other}
$Age\_Decile = 3$	48	45
$Age\_Decile \neq 3$	86	220

Searching for the best score…

- Try ICD9 = x for each value of x
- Try Gender=M, Gender=F
- Try CoarseRegion=NE, =NW, SE, SW..
- Try FineRegion=AA,AB,AC, … DD (4x4 Grid)
- Try Hospital=x, TimeofDay=x, Prodrome=X, …
- [In future… features of cens…

Overfitting Alert!

Example

```
Sat 12-23-2001 (daynum 36882, dayindex 239)
FISHER_PVALUE = 0.000051 RANDOMIZATION_PVALUE = 0.031
35.8% ( 48/134) of today's cases have 30 <= age < 40
17.0% ( 45/265) of   other cases have 30 <= age < 40
```

Table 1: A sample 2x2 Contingency Table

	C_{today}	C_{other}
$Age\_Decile = 3$	48	45
$Age\_Decile \neq 3$	86	220

Multiple component rules

- We would like to be able to find rules like:

  There are a surprisingly large number of children with respiratory problems today

  or

  There are too many skin complaints among people from the affluent neighborhoods

- These are things that would be missed by casual screening
- BUT
  - The danger of overfitting could be much worse
  - It's very computationally demanding
  - How can we be sure the entire rule is meaningful?

Checking two component rules

Table 2: 2x2 Contingency Table 1 for a two component rule

Records from Today matching C_0 and C_1	Records from Other matching C_0 and C_1
Records from Today matching C_1 and differing on C_0	Records from Other matching C_1 and differing on C_0

Table 3: 2x2 Contingency Table 2 for a two component rule

Records from Today matching C_0 and C_1	Records from Other matching C_0 and C_1
Records from Today matching C_0 and differing on C_1	Records from Other matching C_0 and differing on C_1

- Must pass both tests to be allowed to live.

WSARE v2.0

- Inputs:
  1. Date/time-indexed biosurveillance-relevant data stream
  2. Time Window Length
  3. Which attributes to use?
- Outputs:
  1. Here are the records that most surprise me
  2. Here's why
  3. And here's how seriously you should take it

Primary Key	Date	Time	Hospital	ICD9	Prodrome	Gender	Age	Home			Work			Recent Flu Levels	Recent Weather	(Many more...)
								Large Scale	Medium Scale	Fine Scale	Large Scale	Medium Scale	Fine Scale			
h6r32	6/2/2	14:12	Down-town	781	Fever	M	20s	NE	15217	A5	NW	15213	B8	2%	70R	...
t3q15	6/2/2	14:15	River-side	717	Respiratory	M	60s	NE	15222	J3	NE	15222	J3	2%	70R	...
t5hh5	6/2/2	14:15	Smith-field	622	Respiratory	F	80s	SE	15210	K9	SE	15210	K9	2%	70R	...
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

WSARE v2.0

- Inputs:
  1. Date/time-indexed biosurveillance-relevant data stream
  2. Time Window Length
  3. Which attributes to use?
- Outputs:
  1. Here are the records that most surprise me
  2. Here's why
  3. And here's how seriously you should take it

Normally, 8% of cases in the East are over-50s with respiratory problems.

But today it's been 15%

Don't be too impressed!

Taking into account all the patterns I've been searching over, there's a 20% chance I'd have found a rule this dramatic just by chance

WSARE on recent Utah Data

Saturday June 1st in Utah:

The most surprising thing about recent records is:

Normally:

0.8% of records (50/6205) have time before 2pm and prodrome = Hemorrhagic

But recently:

2.1% of records (19/907) have time before 2pm and prodrome = Hemorrhagic

Pvalue = 0.0484042

Which means that in a world where nothing changes we'd

expect to have a result this significant about once

every 20 times we ran the program

Which of the 500 days have Irregularities?

- WARNING: Yet another overfitting opportunity.
- If we took 500 days of randomly generated data, then about 5 days would have a p-value below 0.01
- This can be solved with...
  - A Bonferroni correction
  - The FDR (False Discovery Rate) method [Benjamini & Hochberg, 1995, J. R. Stat Soc, 57 289]

Results on Emergency Dept Data

```
### Rule 1: Tue 05-16-2000 (daynum 36661, dayindex 18)
SCORE = -0.00000000  PVALUE = 0.00000000
 32.84% ( 44/134) of today's cases have Time Of Day4 after 6:00 pm
 90.00% ( 27/ 30) of   other cases have Time Of Day4 after 6:00 pm

### Rule 2: Fri 06-30-2000 (daynum 36706, dayindex 63)
SCORE = -0.00000000  PVALUE = 0.00000000
 19.40% ( 26/134) of today's cases have Place2 = NE and Lat4 = d
  5.71% ( 16/280) of   other cases have Place2 = NE and Lat4 = d

### Rule 3: Wed 09-06-2000 (daynum 36774, dayindex 131)
SCORE = -0.00000000  PVALUE = 0.00000000
 17.16% ( 23/134) of today's cases have Prodrome = Respiratory
and age2 less than 40
  4.53% ( 12/265) of other cases have Prodrome = Respiratory
and age2 less than 40

### Rule 4: Fri 12-01-2000 (daynum 36860, dayindex 217)
SCORE = -0.00000000  PVALUE = 0.00000000
 22.88% ( 27/118) of today's cases have Time Of Day4
after 6:00 pm and Lat2 = s
  8.10% ( 20/247) of   other cases have Time Of Day4
after 6:00 pm and Lat2 = s

### Rule 5: Sat 12-23-2000 (daynum 36882, dayindex 239)
SCORE = -0.00000000  PVALUE = 0.00000000
 18.25% ( 25/137) of today's cases have ICD9 = shortness of breath
and Time Of Day2 before 3:00 pm
  5.12% ( 15/293) of   other cases have ICD9 = shortness of breath
and Time Of Day2 before 3:00 pm

### Rule 6: Fri 09-14-2001 (daynum 37147, dayindex 504)
SCORE = -0.00000000  PVALUE = 0.00000000
 66.67% ( 30/ 45) of today's cases have Time Of Day4 before 10:00 am
 18.42% ( 42/228) of   other cases have Time Of Day4 before 10:00 am
```


Sanity check

- What happens if we generate a fake database in which we know that there can be no relation between date and case features?
- This can be achieved by shuffling all the dates in the database.
- The days detected by FDR are then...

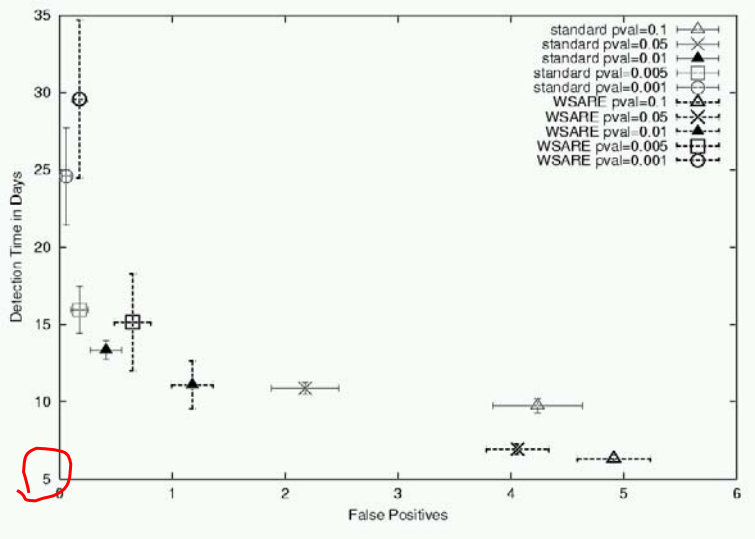
Survived Randomized

- No days reported as abnormal

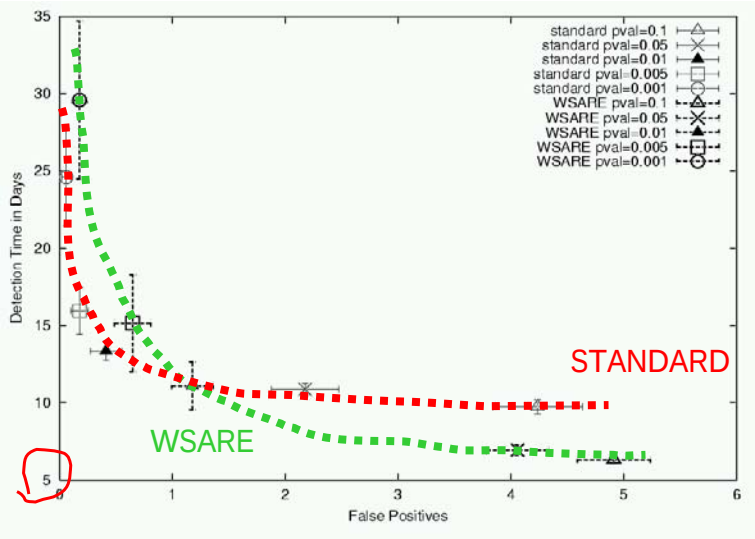
Simulation-based validation

- We need to understand sensitivity and specificity.
- Have developed a very fast simulator in which people go to work, go out for food, live in families.
- Background of mild diseases: some food-borne, some occupational, some contagious etc.
- A nasty disease is introduced at some point.
  - Can WSARE detect it?
  - How does WSARE compare to a standard sensible epidemiological alternative?

WSARE Simulation Sensitivity vs Specificity on GRIDSIM



WSARE Simulation Sensitivity vs Specificity on GRIDSIM

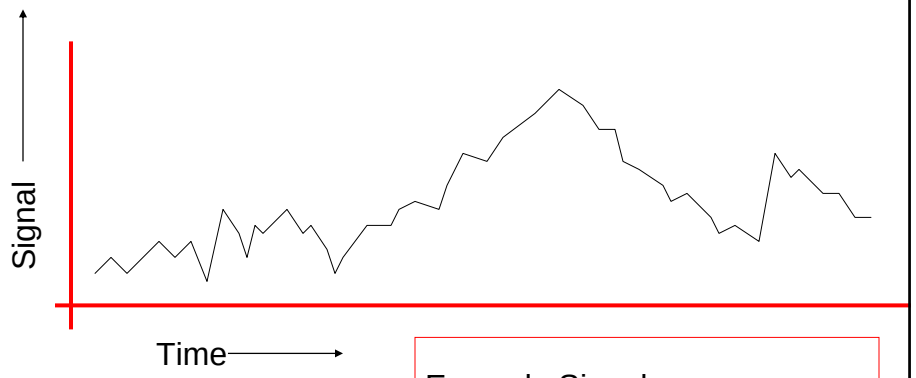


WSARE 3.0

- "Taking into account recent flu levels..."
- "Taking into account that today is a public holday..."
- "Taking into account that this is Spring..."
- "Taking into account recent heatwave..."
- "Taking into account that there's a known natural Food-borne outbreak in progress..."

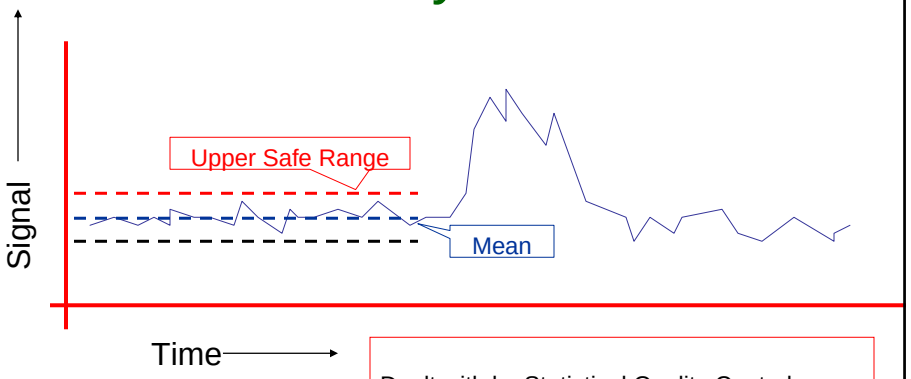
Bonus: More efficient use of historical data

Conditioning on observed environment: Well understood for Univariate Time Series



Example Signals:
- Number of ED visits today
- Number of ED visits this hour
- Number of Respiratory Cases Today
- School absenteeism today
- Nyquil Sales today

An easy case

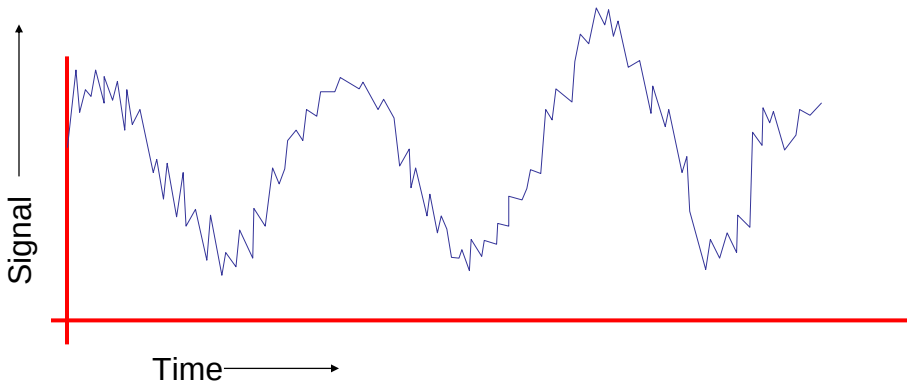


Dealt with by Statistical Quality Control

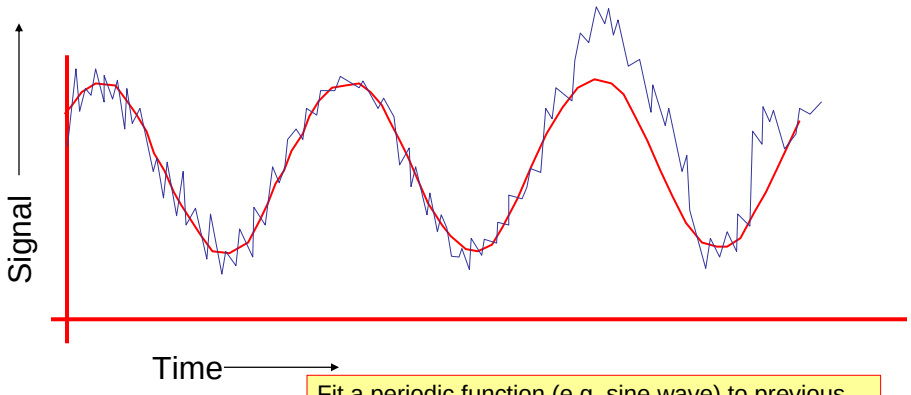
Record the mean and standard deviation up the the current time.

Signal an alarm if we go outside 3 sigmas

Conditioning on Seasonal Effects



Seasonal Effects

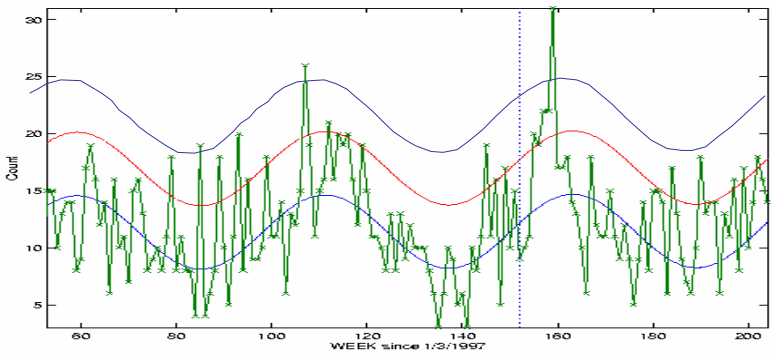


Fit a periodic function (e.g. sine wave) to previous data. Predict today's signal and 3-sigma confidence intervals. Signal an alarm if we're off.

Reduces False alarms from Natural outbreaks.

Different times of year deserve different thresholds.

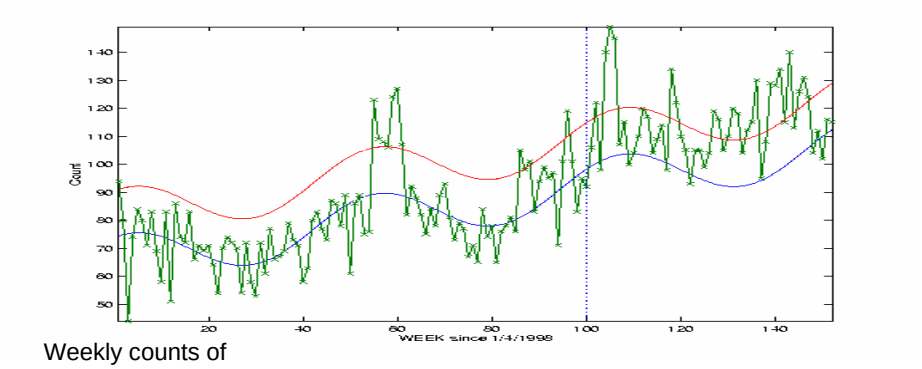
Example [Tsui et. Al]



Weekly counts of P&I from week 1/98 to 48/00

From: *"Value of ICD-9–Coded Chief Complaints for Detection of Epidemics"*, Fu-Chiang Tsui, Michael M. Wagner, Virginia Dato, Chung-Chou Ho Chang, AMIA 2000

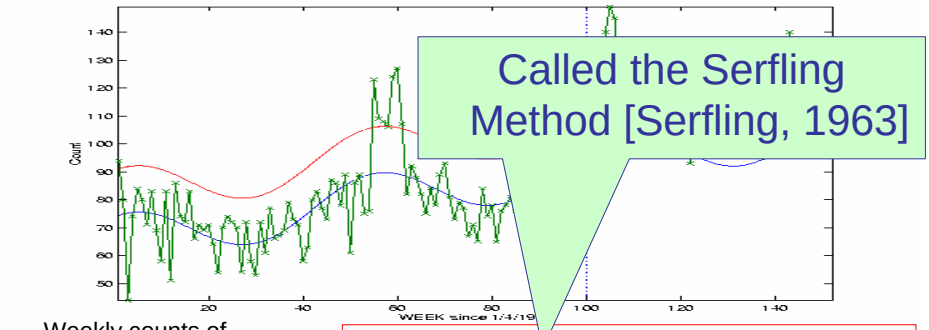
Seasonal Effects with Long-Term Trend



Weekly counts of IS from week 1/98 to 48/00.

From: *"Value of ICD-9–Coded Chief Complaints for Detection of Epidemics"*, Fu-Chiang Tsui, Michael M. Wagner, Virginia Dato, Chung-Chou Ho Chang, AMIA 2000

Seasonal Effects with Long-Term Trend



Called the Serfling Method [Serfling, 1963]

Weekly counts of IS from week 1/98 to 48/00.

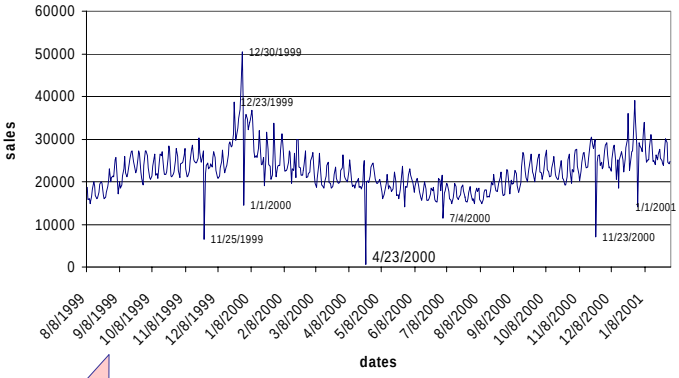
From: *"Value of ICD-9–Coded Chief Complaints for Detection of Epidemics"*, Fu-Chiang Tsui, Michael M. Wagner, Virginia Dato, Chung-Chou Ho Chang, AMIA 2000

Fit a periodic function (e.g. sine wave) plus a linear trend:

E[Signal] = a + bt + c sin(d + t/365)

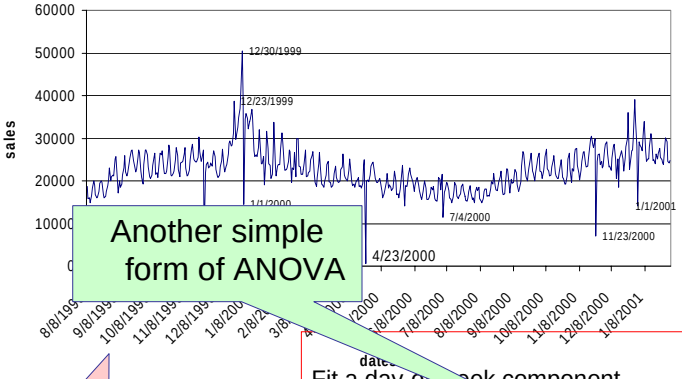
Good if there's a long term trend in the disease or the population.

Day-of-week effects



From: *"Using Grocery Sales Data for the Detection of Bio-Terrorist Attacks"*, Goldenberg, Shmueli and Caruana, 2002

Day-of-week effects



Another simple form of ANOVA

From: *"Using Grocery Sales Data for the Detection of Bio-Terrorist Attacks"*, Goldenberg, Shmueli and Caruana, 2002

Fit a day-of-week component

$E[Signal] = a + delta_{day}$

E.G: $delta_{mon}$= +5.42, $delta_{tue}$= +2.20, $delta_{wed}$= +3.33, $delta_{thu}$= +3.10, $delta_{fri}$= +4.02, $delta_{sat}$= -12.2, $delta_{sun}$= -23.42

Analysis of variance

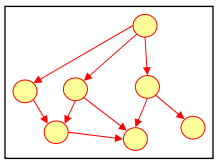
- Good news:

  If you're tracking a daily aggregate (e.g. number of flu cases in your ED, or Nyquil Sales)...then ANOVA can take care of many of these effects.

- But...

  What if you're tracking a whole joint distribution of transactional events?

Idea: Bayesian Networks



"Patients from West Park Hospital are less likely to be young"

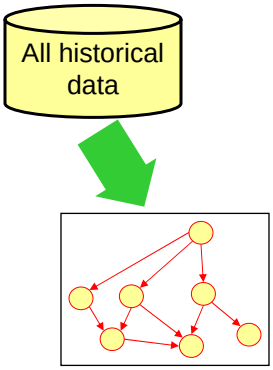
"On Cold Tuesday Mornings the folks coming in from the North part of the city are more likely to have respiratory problems"

"The Viral prodrome is more likely to co-occur with a Rash prodrome than Botulinic"

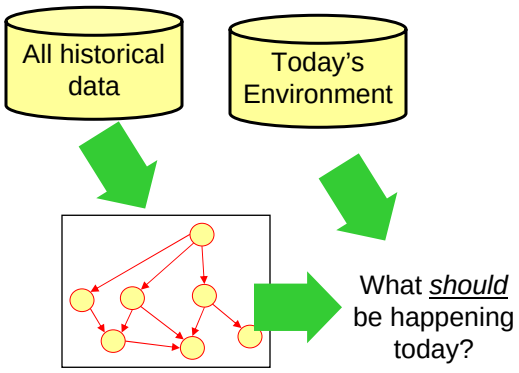
"On the day after a major holiday, expect a boost in the morning followed by a lull in the afternoon"

WSARE 3.0

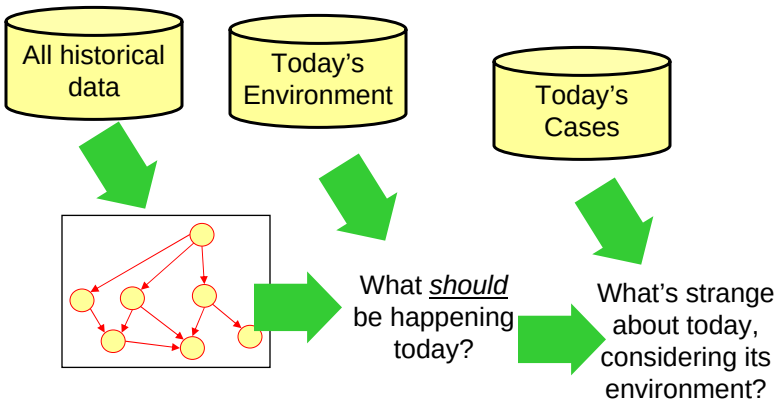
WSARE 3.0



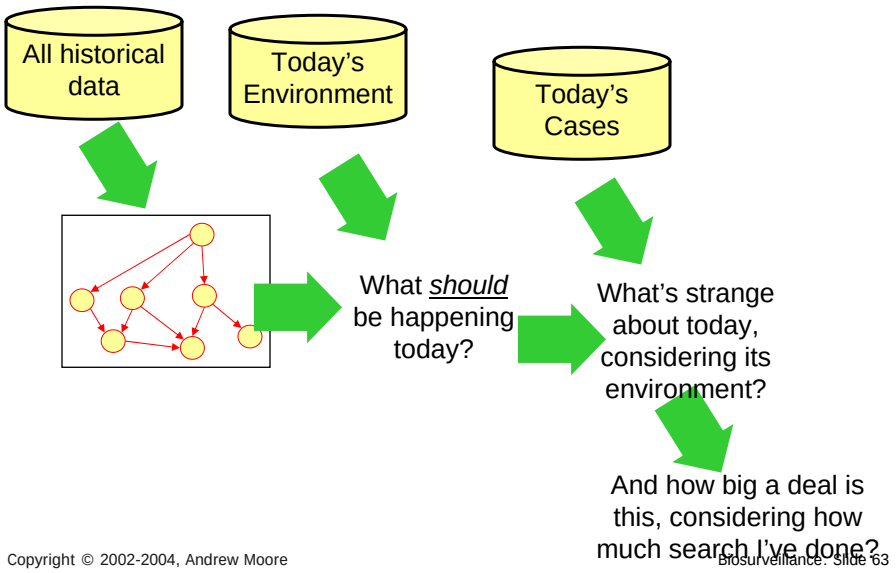
WSARE 3.0



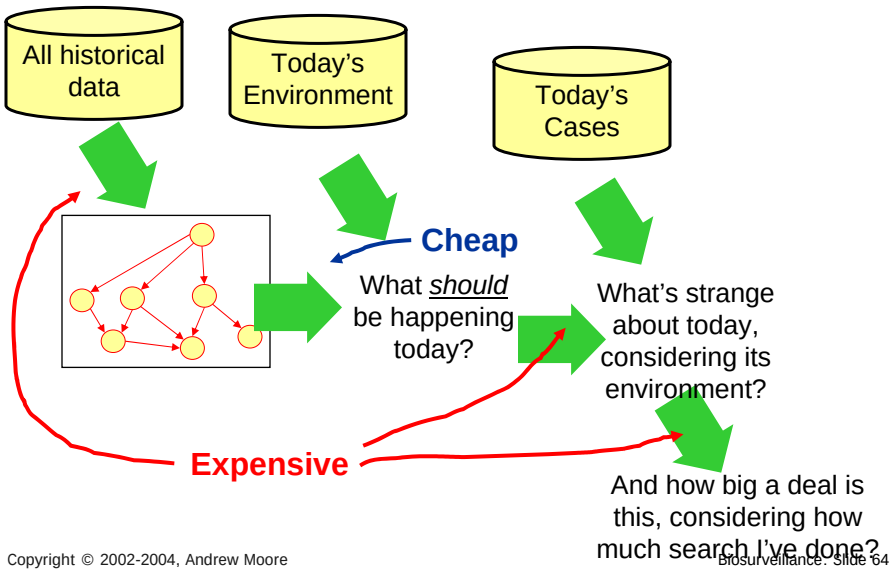
WSARE 3.0

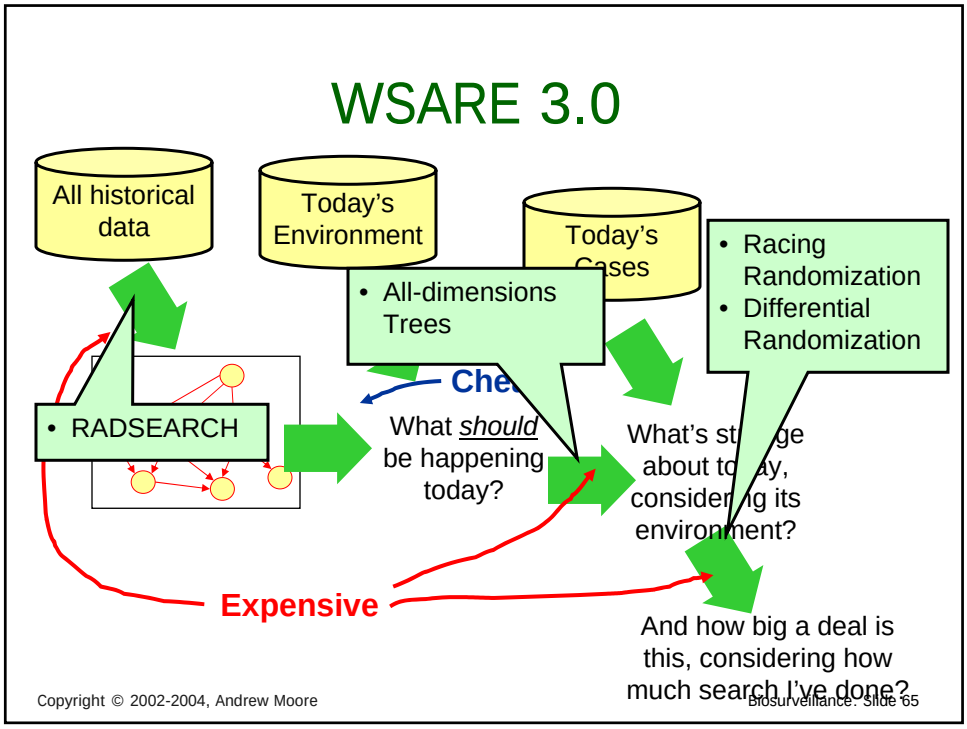


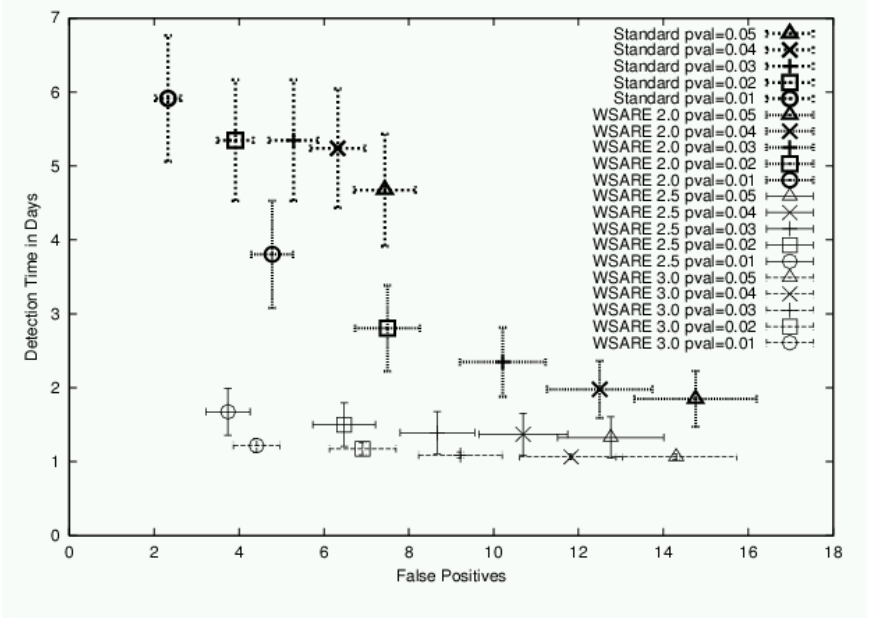
WSARE 3.0



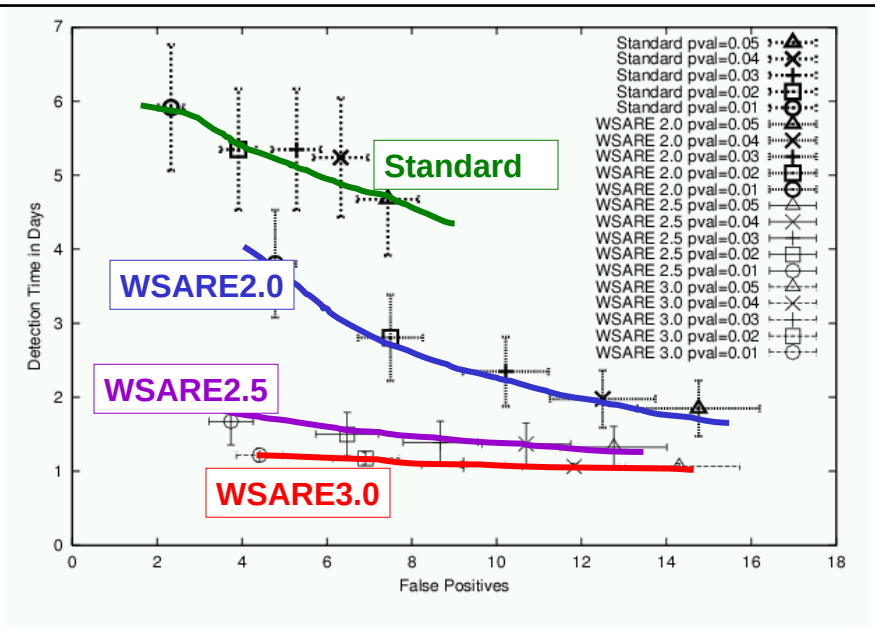
WSARE 3.0







Results on Simulation



Results on Simulation

Conclusion

- One approach to biosurveillance: one algorithm monitoring millions of signals derived from multivariate data

  instead of

  Hundreds of univariate detectors

- Modeling historical data with Bayesian Networks to allow conditioning on unique features of today
- Computationally intense unless we're tricksy!

- Searching over thousands of contingency tables on a large database…
- …only we have to do it 10,000 times on the replicas during randomization
- …we also need to learn Bayes Nets from databases with millions of records…
- …and keep relearning them as data arrives online…
- …in the end we typically search about a billion alternative Bayes net structures for modeling 800,000 records in 10 minutes

allow conditioning unique features of today
- Computationally intense unless we're tricksy!

Conclusion

- One approach to biosurveillance: one algorithm monitoring millions of signals derived from multivariate data

  instead of

  Hundreds of univariate detectors

- Modeling historical data with Bayesian Networks to allow conditioning on unique features of today
- Computationally intense unless we're tricksy!
- WSARE 2.0 Deployed during the past year
- WSARE 3.0 about to go online
- WSARE now being extended to additionally exploit over the counter medicine sales