Influenza disease outbreak detection using multiple electronic data sources

Sharon K. Greene,[1] Martin Kulldorff,[1] Jie Huang,[2]
Richard Brand,[3] Ken Kleinman,[1] John Hsu,[2,3] and Richard Platt[1]

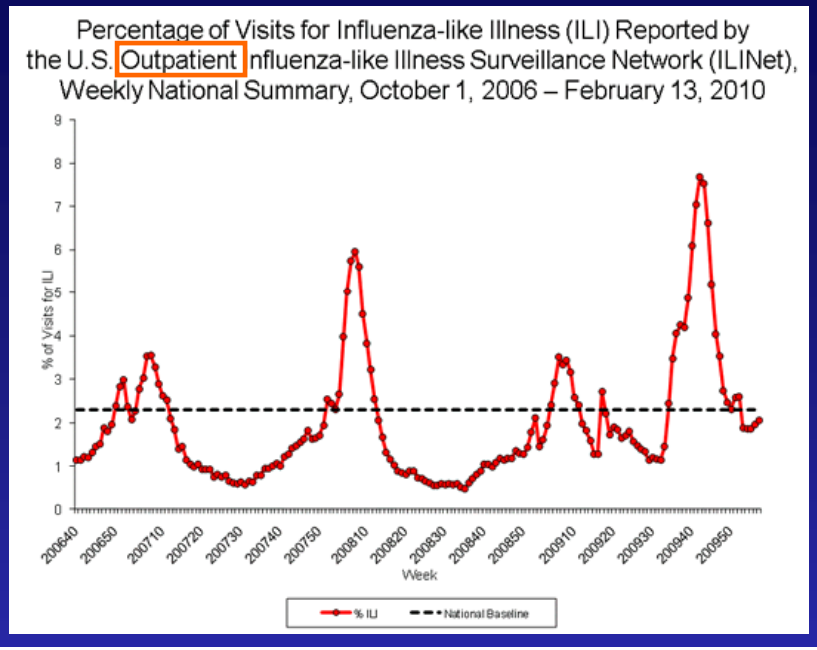
[1] Dept of Population Medicine, Harvard Medical School &
Harvard Pilgrim Health Care Institute, Boston, MA

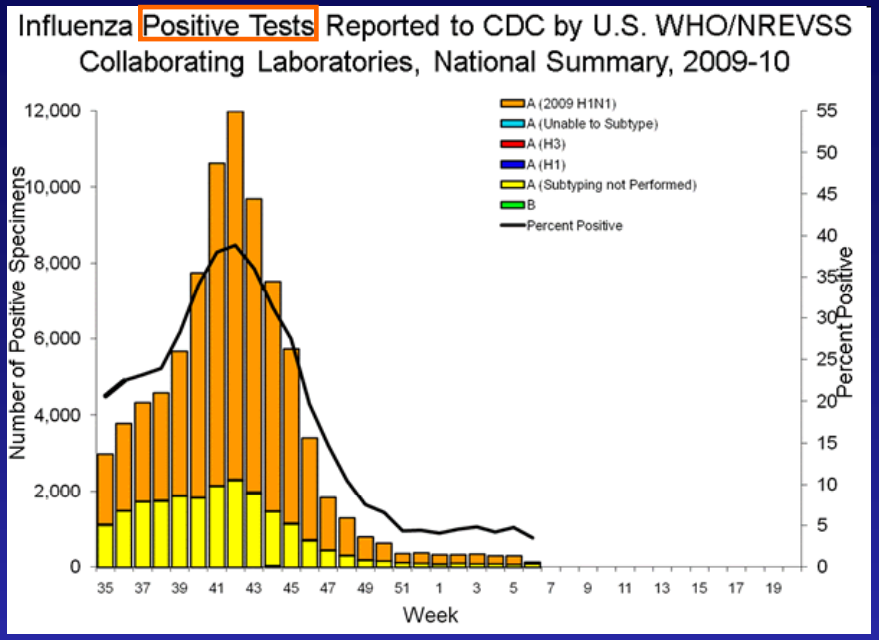
[2] Center for Health Policy Studies, Division of Research,
Kaiser Permanente Medical Care Program, Oakland, CA

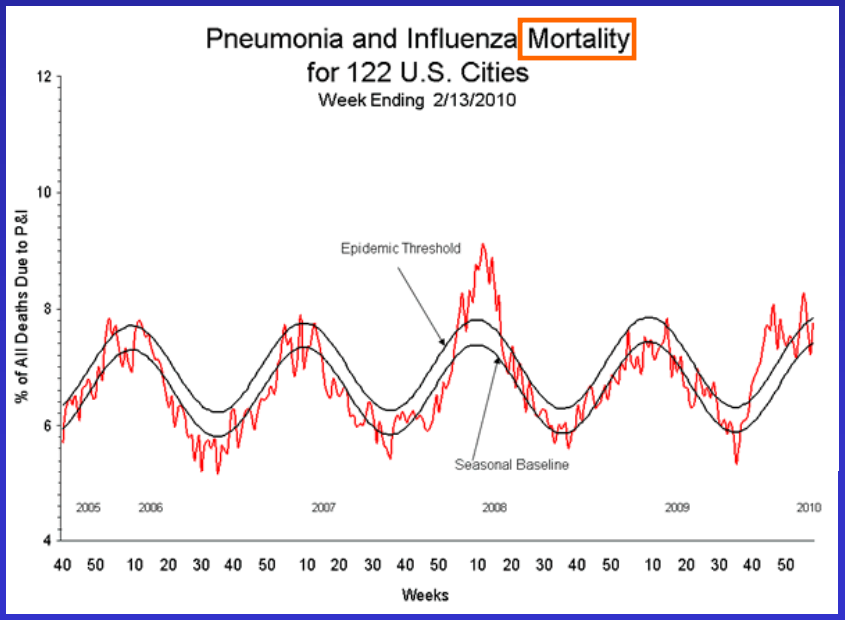
[3] Dept of Epidemiology and Biostatistics,
Univ. of California, San Francisco, CA

Research support provided by the
National Institute of General Medical Sciences under the
Models of Infectious Disease Agent Study (MIDAS) program

Traditional Influenza Surveillance







Disadvantages:

- Sentinel sites only
- Weekly (timeliness)
- Purely temporal (infection not uniform everywhere)
- Requires active reporting

www.cdc.gov/flu/weekly/

Possible Electronic Data Sources for Infectious Disease Surveillance

Before seeking care

- Internet search
- Telephone triage service call
- Over-the-counter medication purchase

Seeking care

- Ambulatory care (AC) visit
- Emergency department (ED) visit
- House calls

Resulting from seeking care

- Lab test ordered
- Lab test result
- Medication prescribed
- Medication dispensed
- Hospital admission
- Hospital discharge

Other

- Ambulance dispatched
- Death

Which stream(s) most useful for identifying localized excess activity?

No Single Data Stream Optimal

Stream	Advantage	Disadvantages
RT-PCR tests	Specific	Non-representative: ordered at clinician discretion, disproportionately at beginning of season and for high-risk patients
AC, ED	Sensitive	Non-specific, non-representative: ruling out serious illness, care-seeking influenced by media
Requiring fever in AC, ED definition	Specific	Less sensitive
Antiviral medications	Specific	Non-representative Treatment or prophylaxis? Amantadine used for Parkinson's dx
Hospital admissions	Timely indicator of severe illness	Small sample size
Hospital discharges	More specific than admissions	Less timely; small sample size

Study Population

- Integrated health care delivery system
- 3.3 million members
- Central Valley and San Francisco Bay
- Of 18 medical centers, 12 electronically recorded patient temperature

Available Data Sources

Before seeking care

- Internet search
- Telephone triage service call
- Over-the-counter medication purchase

Seeking care

- Ambulatory care (AC) visit
- Emergency department (ED) visit
- House calls

Resulting from seeking care

- Lab test ordered
- Lab test result
- Medication prescribed
- Medication dispensed
- Hospital admission
- Hospital discharge

Other

- Ambulance dispatched
- Death

Research Article

Statistics in Medicine

Received 22 January 2010, Accepted 25 January 2010 Published online in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3883

Timely detection of localized excess influenza activity in Northern California across patient care, prescription, and laboratory data

Sharon K. Greene,[a]*† Martin Kulldorff,[a] Jie Huang,[b] Richard J. Brand,[c] Kenneth P. Kleinman,[a] John Hsu[b,c] and Richard Platt[a]

Timely detection of clusters of localized influenza activity in excess of background seasonal levels could improve situational awareness for public health officials and health systems. However, no single data type may capture influenza activity with optimal sensitivity, specificity, and timeliness, and it is unknown which data types could be most useful for surveillance. We compared the performance of 10 types of electronic clinical data for timely detection of influenza clusters throughout the 2007/08 influenza season in northern California. Kaiser Permanente Northern California generated zip code-specific daily episode counts for: influenza-like illness (ILI) diagnoses in ambulatory care (AC) and emergency departments (ED), both with and without regard to fever; hospital admissions and discharges for pneumonia and influenza; antiviral drugs dispensed (Rx); influenza laboratory tests ordered (Tests); and tests positive for influenza type A (FluA) and type B (FluB). Four credible events of localized excess illness were identified. Prospective surveillance was mimicked within each data stream using a space–time permutation scan statistic, analyzing only data available as of each day, to evaluate the ability and timeliness to detect the credible events. AC without fever and Tests signaled during all four events and, along with Rx, had the most timely signals. FluA had less timely signals. ED, hospitalizations, and FluB did not signal reliably. When fever was included in the ILI definition, signals were either delayed or missed. Although limited to one health plan, location, and year, these results can inform the choice of data streams for public health surveillance of influenza.

Keywords: influenza; outbreak detection; spatio-temporal analysis

Available Data Streams

Data stream	Explanation	Temporal data element
AC±F	Influenza-like illness in ambulatory care	Encounter
AC+F	" with fever	
ED±F	Influenza-like illness in emergency department	
ED+F	" with fever	

Condition	ICD-9 code(s)
Viral infection	079.3, 079.89, 079.99
Acute pharyngitis	460, 462
Acute laryngitis and tracheitis	464.0, 464.1, 464.2, 465
Acute bronchitis and bronchiolitis	466.0, 466.19
Other diseases of the upper respiratory tract	478.9
Pneumonia	480.8, 480.9, 481, 482.40, 482.41, 482.49, 484.8, 485, 486
Influenza	487
Throat pain	784.1
Cough	786.2

Available Data Streams

Data stream	Explanation	Temporal data element
AC±F	Influenza-like illness in ambulatory care	Encounter
AC+F	" with fever	
ED±F	Influenza-like illness in emergency department	
ED+F	" with fever	
Admissions	Pneumonia or influenza hospital inpatient admissions (text strings)	Admission
Discharges	Same as above, but with a primary discharge diagnosis of pneumonia or influenza (ICD-9 codes)	

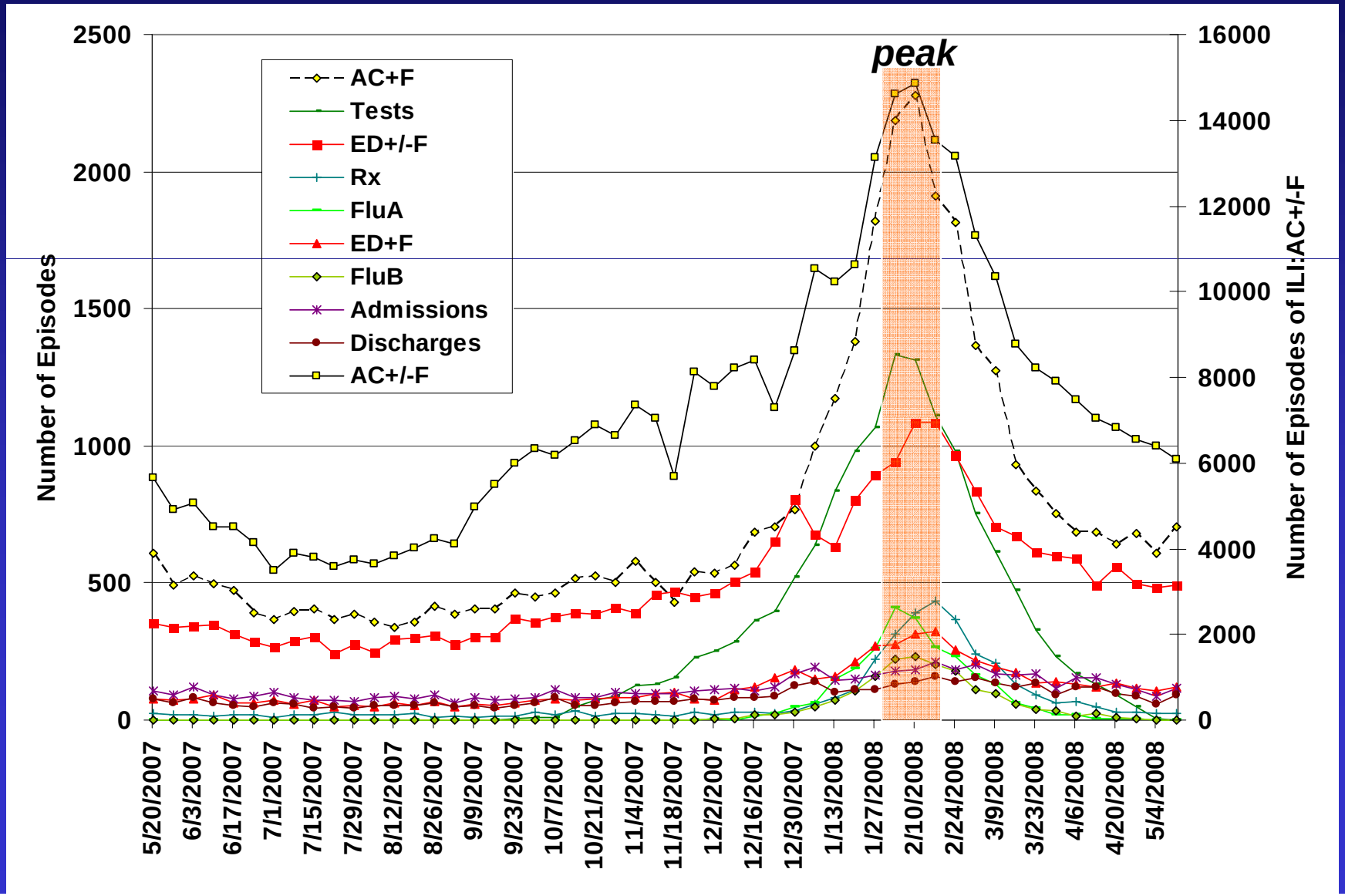
Available Data Streams

Data stream	Explanation	Temporal data element
AC±F	Influenza-like illness in ambulatory care	Encounter
AC+F	" with fever	
ED±F	Influenza-like illness in emergency department	
ED+F	" with fever	
Admissions	Pneumonia or influenza hospital inpatient admissions (text strings)	Admission
Discharges	Same as above, but with a primary discharge diagnosis of pneumonia or influenza (ICD-9 codes)	
Rx	Antivirals (amantadine and oseltamivir)	Dispensing

Available Data Streams

Data stream	Explanation	Temporal data element
AC±F	Influenza-like illness in ambulatory care	Encounter
AC+F	" with fever	
ED±F	Influenza-like illness in emergency department	
ED+F	" with fever	
Admissions	Pneumonia or influenza hospital inpatient admissions (text strings)	Admission
Discharges	Same as above, but with a primary discharge diagnosis of pneumonia or influenza (ICD-9 codes)	
Rx	Antivirals (amantadine and oseltamivir)	Dispensing
Tests	RT-PCR tests ordered	Specimen collected
FluA	RT-PCR tests positive for influenza type A	
FluB	" B	

Weekly episodes of influenza-associated data streams, 5/20/07-5/17/08



Space-Time Permutation Scan Statistic: Properties

- Adjusts for purely geographical or purely temporal variation
- Simultaneously tests for outbreaks of any size at any location
- Accounts for multiple testing

SaTScan™ (www.satscan.org)

- Surveillance period: 9/30/07-5/17/08
- Historical period: 5/20/07-9/29/07
- Episodes (i.e., repeat visits excluded)

Space-Time Permutation Scan Statistic: Analysis Options

- Retrospective
  - Once
  - Using all data available
  - Identify most unusual clusters during study period
- Prospective
  - Daily, for early detection
  - Using data available as of each surveillance day
  - Identify most unusual cluster each day

Credible Retrospective Influenza Cluster #1: Centroid in Bay Area

Data Stream	Cluster Duration	# of zip codes	Obs	Exp	Obs/ Exp	p-value
AC+F	12/7-1/3	58	710	557	1.3	0.0003

Credible Retrospective Influenza Cluster #1: Centroid in Bay Area

Data Stream	Cluster Duration	# of zip codes	Obs	Exp	Obs/ Exp	p-value
AC+F	12/7-1/3	58	710	557	1.3	0.0003
Tests	12/8-1/2	94	506	385	1.3	0.0003
AC±F	12/15-1/5	19	1,620	1,387	1.2	0.001

Credible Retrospective Influenza Cluster #1: Centroid in Bay Area

Data Stream	Cluster Duration	# of zip codes	Obs	Exp	Obs/ Exp	p-value
AC+F	12/7-1/3	58	710	557	1.3	0.0003
Tests	12/8-1/2	94	506	385	1.3	0.0003
AC±F	12/15-1/5	19	1,620	1,387	1.2	0.001
ED+F	12/11-12/25	10	13	3	4.79	n.s.
ED±F	12/16-12/26	69	313	248	1.3	n.s.
FluA	12/17-12/30	9	6	0.81	7.4	n.s.

Credible Retrospective Cluster Characteristics

- Detection at $p<0.005$ by ≥3 streams
- Excess risk in ≥4 streams
- Occurrence when >5% specimens statewide tested positive

Four Credible Retrospective Clusters

#	Centroid Location	Duration
1	Bay Area	Dec-Jan
2	Fresno	Jan-Feb
3	Sacramento	Feb
4	Bay Area	Mar-Apr

Prospective Detection of Cluster #1

	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA

Prospective Detection of Cluster #1

	# of days after 1st signal that each data stream signaled						
	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA

Prospective Detection of Cluster #1

	# of days after 1st signal that each data stream signaled						
Min. RI (years)	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA
1							
2							
5							
10							
25							

Prospective Detection of Cluster #1

	# of days after 1st signal that each data stream signaled						
Min. RI (years)	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA
1	-	0	-	-	-	0	-
2	-	0	-	-	-	-	-
5	-	2	-	-	-	-	-
10	-	14	-	-	-	-	-
25	-	14	-	-	-	-	-

Prospective Detection of Cluster #2

	# of days after 1st signal that each data stream signaled						
Min. RI (years)	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA
1	-	28	-	-	0	4	-
2	-	28	-	-	0	4	-
5	-	28	-	-	5	4	-
10	-	28	-	-	5	4	-
25	-	28	-	-	5	4	-

Prospective Detection of Cluster #3

	# of days after 1st signal that each data stream signaled						
Min. RI (years)	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA
1	2	0	21	8	21	6	13
2	2	0	22	8	-	6	14
5	4	1	22	9	-	7	-
10	4	1	22	9	-	7	-
25	4	2	22	9	-	7	-

Prospective Detection of Cluster #4

	# of days after 1st signal that each data stream signaled						
Min. RI (years)	AC+F	AC±F	ED+F	ED±F	Rx	Tests	FluA
1	-	24	-	41	-	0	4
2	-	25	-	-	-	-	10
5	-	25	-	-	-	-	10
10	-	25	-	-	-	-	11
25	-	25	-	-	-	-	11

Summary of Prospective Analyses

- AC±F and Tests signaled during all 4 clusters
  - Most timely for 3 clusters
- Rx most timely for one cluster
- FluA also had timely signals
- Not reliable: ED+F, ED±F, hospital admissions and discharges, and FluB
- When fever included in ILI definition for AC or ED, signals were delayed or missed

Limitations

- No external gold standard for credible influenza activity
- One health care system, one state, one influenza season

Available Data Sources

Before seeking care

- Internet search
- Telephone triage service call
- Over-the-counter medication purchase

Seeking care

- Ambulatory care (AC) visit
- Emergency department (ED) visit
- House calls

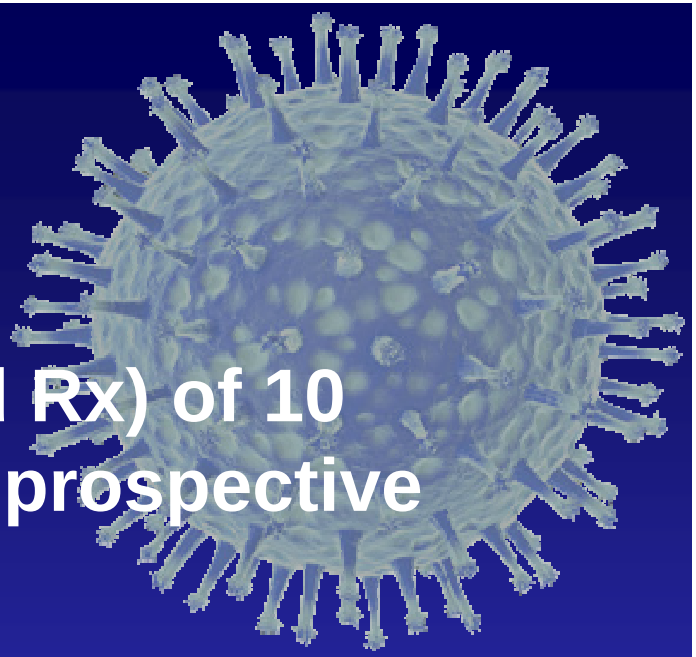
Resulting from seeking care

- Lab test ordered
- Lab test result
- Medication prescribed
- Medication dispensed
- Hospital admission
- Hospital discharge

Other

- Ambulance dispatched
- Death

Conclusions



- Using only 3 (AC±F, Tests, and Rx) of 10 available data streams, timely prospective detection feasible
- AC streams more useful than ED streams
- Inform selection and development of data for enhanced public health surveillance
  - Univariate or multivariate methods
- Additional research needed to confirm which data streams most promising

Limitations

- Space-time clusters may occur for other reasons than disease outbreaks
- Automated detection systems does not replace the observant eyes of physicians and other health workers.
- Epidemiological investigations by physicians, epidemiologists or microbiologists are needed to confirm or dismiss the signals

SaTScan Software

Free. Download from www.satscan.org

Registered users in 156 countries:

1. USA
2. Canada
3. United Kingdom
4. Brazil
5. Italy

. . .

100s. Albania, Bhutan, Burma, Fiji, Grenada, Guinea, Iraq, Macao, Madagascar, Malawi, Malta, etc

Acknowledgement

Research funded by:

Alfred P Sloan Foundation

Centers for Disease Control and Prevention

Massachusetts Department of Health

National Cancer Institute

National Institute of Child Health and Development

National Institute of General Medical Sciences:

Modeling Infectious Disease Agent Study (MIDAS)