Syndromic Surveillance Using Electronic Health Record Clinical Data from Hospital and Ambulatory Settings: An Annotated Bibliography

A Report from the International Society for Disease Surveillance

April 2013

For information about this report, contact:
Becky Zwickl, MPH
ISDS Policy Analyst
617-779-0880
bzwickl@syndromic.org

International Society for Disease Surveillance

Syndromic Surveillance Using Electronic Health Record Clinical Data from Hospital and Ambulatory Settings: An Annotated Bibliography

Table of Contents

45. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions.

46. Ambulatory-care diagnoses as potential indicators of outbreaks of gastrointestinal illness—Minnesota.

Purpose:

The purpose of this annotated bibliography is to summarize published articles and abstracts on the use of hospital (both inpatient and emergency department) and ambulatory electronic health records (EHR) data for syndromic surveillance. This bibliography adds to the growing evidence base for the feasibility and utility of utilizing hospital and ambulatory EHR records for syndromic surveillance.

How to Use this Bibliography:

This bibliography is designed to be a resource for both researchers and practitioners. Researchers may find the information useful when conducting literature searches on syndromic surveillance. Practitioners may use these articles as a basis for researching and establishing best practices in syndromic surveillance using data from clinical settings. This document currently contains articles retrieved as of April 2013. Additional articles may be added in the future as they are identified and deemed relevant.

The articles are listed by topic area (as seen in the table of contents). These topic areas are: Hospital: Inpatient and Emergency Department (ED); Hospital and Ambulatory; Ambulatory; and Potential Uses Utilizing Future Data Elements. A short description at the start of each section summarizes the purpose of each type of clinical data source.

Within each topic area, articles are arranged alphabetically by last name of the paper's first author. You may also find it useful to do a keyword search using the find (CTRL+f) function on your keyboard.

Process:

Article search:

Articles and abstracts were extracted from a variety of sources using keyword searches of databases (listed below), as well as manual searches. These manual searches included: a search through ISDS conference abstracts and a circular search through the article references for relative citations.

- Databases/techniques used:
  - Google scholar
  - Pub Med
  - EBSCO
  - Direct referral to articles by stakeholders, Meaningful Use Workgroup members
  - Manual search through ISDS conference proceedings
  - Used citations from articles retrieved to find related articles

- Keyword search terms:
  - Inpatient electronic health record
  - Ambulatory electronic health record
  - Ambulatory (care) syndromic surveillance
  - Inpatient syndromic surveillance
  - Hospital syndromic surveillance
  - Outpatient syndromic surveillance
  - Inpatient EHR syndromic surveillance
  - Ambulatory EHR syndromic surveillance
  - Chronic disease syndromic surveillance

Outline of Data Gathered:
With each article, the following information was extracted and summarized.

- *Data source:* Ambulatory, emergency department, hospital inpatient, or some combination of the three
- *Objectives of project:* The purpose of the article or abstract
- *Data extracted from records:* What data elements were utilized?
- *Case definitions (if applicable):* What case definitions were applied (i.e., how did the authors define influenza-like illness?)
- *Data use:* Uses for public health, such as situational awareness or outbreak detection
- *Data analysis:* What analytic methods were used?
- *Statistics/Results:* What were the results both qualitative and quantitative?
- *Data quality:* Information on timeliness, sensitivity, specificity, completeness, etc.

Hospital: Inpatient and Emergency Department (ED)

This section, *Hospital: Inpatient and Emergency Department (ED),* describes research and practice in using data from hospital EHR records for syndromic surveillance purposes. In some cases data is gathered from inpatient settings, while in others it is gathered from EDs only, and in the remaining cases EHR information is gathered from both inpatient and ED sources.

1. ***Title:***
   An Emergency Department Based Syndromic Surveillance System for Meningitis and Encephalitis, Maricopa County, AZ, 2004

   Citation:
   Arboleda N, Fleischauer AT, Sejvar J, Diggs A, Schumacher M, Santana S, Engelthaler D, Komatsu K, Hughes S, Jones G, Hutwagner L. An Emergency Department Based Syndromic Surveillance System for Meningitis and Encephalitis, Maricopa County, AZ, 2004. *Advances in Disease Surveillance* 2006;1:4.

   Summary:
   - *Data source:* **Emergency Department** chief complaint data
     - ED records were gathered from 948 patients with verified M/E cases
     - Data from Maricopa County, Arizona
   - *Objectives of project:*
     - To evaluate the syndrome definitions developed for meningitis and encephalitis (M/E)
     - To see if ED data could provide timely reports on/detect M/E outbreaks
   - *Data extracted from records:*
     - ED visit date/time
     - Age
     - Gender
     - Chief complaint
     - Discharge diagnosis
   - *Case definitions (if applicable):*
     - 5 increasingly specific case definitions were created
     - Each of the 5 case definitions for M/E were evaluated
     - All but the final, most restrictive, case definition increased noise without detecting aberration events
     - Final case definition used: 18% sensitivity, 99% specificity at detecting M/E; produce signal before increase in M/E reports was noted
   - *Data use:* Outbreak detection
   - *Data analysis:* Data were analyzed to determine the best balance of positive predictive value, sensitivity, and specificity (focus mainly on positive predictive value to avoid excessive noise)
   - *Statistics:* 570 (70%) of Maricopa County M/E cases for study time frame were analyzed to see how they fit with the 5 possible case definitions
   - *Data quality (e.g., timeliness, accuracy, validity, completion): N/A*

2. ***Title:***
Rapid Identification of Pneumonias in BioSense Data Using Radiology Text Reports.

Citation:
Asatryan A, Ma H, English R, Tokars J. Rapid Identification of Pneumonias in BioSense Data Using Radiology Text Reports. *Advances in Disease Surveillance.* 2007;4:82.

Summary:
- *Data source:* BioSense data gathered from Feb 2006-Jan 2007
  - Data all extracted from **hospital visits**
  - Not specified if this is inpatient only
  - Geographic region not specified
- *Objectives of project:* To see if BioSense data can be effectively scanned for pneumonia utilizing a keyword search
- *Data extracted from records:*
  - Chest x-ray reports (67,714)
  - Resulting ICD-9 final diagnosis codes
- *Case definitions (if applicable):*
  - SAS program was set up to find the following words in x-ray reports:
    - Airspace
    - Consolidation
    - Density
    - Infiltrate
    - Opacity
    - Pneumonia/pneumonitis
- *Data use:* Outbreak detection/inform health services
- *Data analysis:*
  - 400 sample records were selected
    - 300 contained pneumonia-related keywords, 100 did not
  - 400 sample records were analyzed by physicians: 260 (out of the true 100) of these records were determined to have none of the keywords
  - Keyword results were analyzed for their association (Odds Ratio) with an ICD-9 diagnosis code of pneumonia
- *Statistics:*
  - 5 of the keywords had a statistically significant correlation (Odds Ratio > 1.0) with the presence of a final diagnosis of pneumonia
  - SAS parsing program showed 98.5% sensitivity and 98.6% specificity for identifying keywords
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: radiology reports are usually available 1-2 days after the radiology is ordered
  - 98.5% sensitivity, 98.6% specificity
  - Accuracy: One or more of 5 keywords was found in 71.7% of patients with pneumonia diagnosis, only found in 30.8% of patients without pneumonia diagnosis

3. ***Title:***
Syndromic surveillance in public health practice, New York City.

Citation:
Heffernan R, Mostashari F, Das D, et al. Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases* [Serial on the internet], 2004 May. Available at: http://www.cdc.gov/ncidod/EID/vol10no5/03-0646.htm. Accessed: May 10, 2008.

Summary:

- *Data source:* Health record information from **emergency departments** in New York City
- *Objectives of project:* To present the findings from the first year of an ED electronic health record syndromic surveillance system
- *Data extracted from records:*
    - Visit date and time
    - Age (in years)
    - Sex
    - Home zip code
    - Free-text chief complaint
- *Case definitions (if applicable):*
    - Common cold (nasal drip, congestion, stuffiness) exclude: chest congestion, sore throat
    - Sepsis (sepsis, cardiac arrest, unresponsive, unconscious, dead on arrival)
    - Respiratory (cough, shortness of breath, difficulty breathing, croup, dyspenea, bronchitis, pneumonia, hypoxia, upper respiratory illness, chest congestion) exclude: cold
    - Diarrhea (diarrhea, enteritis, gastroenteritis, stomach virus)
    - Fever (fever, chills, flu, viral syndrome, body ache and pain, malaise) exclude: hay fever
    - Rash (vesicles, chicken pox, folliculitis, herpes, shingles) exclude: thrush, diaper and genital rash
    - Asthma (asthma, wheezing, reactive airway, chronic obstructive airway disease)
    - Vomiting (vomiting, food poisoning)
- *Data use:* Information for health services/outbreak detection/situational awareness
- *Data analysis:*
    - Data were sorted by syndrome (definition not presented)
    - Chief complaints were assigned to a single syndrome grouping based on a pre-determined hierarchy
    - Each syndrome category was analyzed separately, looking for moderate-large scale events
    - Calculated expected city-wide amounts and compared to actual counts
        - Spatial
        - Temporal
- *Statistics/Results:*
    - Average number of visits per day (citywide): 6,780
    - 64% of respiratory syndrome signals, 95% of fever signals coincided with influenza outbreaks in the city
    - 83% of diarrhea signals and 88% of vomiting signals coincided with known outbreaks

    - Respiratory: no known outbreaks detected by other surveillance systems, no significant signals detected by syndromic surveillance
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
    - Timeliness: data are transmitted every 24 hours (for encounters occurring from midnight to midnight)
        - Analysis occurs daily (data should be received from each ED by 10 am, analysis is usually complete by 1 pm)
    - Timeliness: Best detects more recent signals (1-3 days of additional syndrome) rather than longer-running outbreaks (7+ days)
    - Completeness: Median percentage of EDs sending data for analysis was 95%, range: 63% to 100%
    - Completeness: chief complaint field blank/not useful in 4.1% of records
    - Validity: High levels of agreement between known flu outbreaks and detection of influenza syndrome AND known GI outbreaks and detection of diarrhea/vomiting

---

4. ***Title:***
Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance.

Citation:
Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. *Proceedings AMIA Symposium*. 2002:345-349.

Summary:
- *Data source:* **Emergency department** data
    - Data used from January-December 2000
    - Data retrieved from University of Pittsburgh Medical Center
- *Objectives of project:* To determine the accuracy of 3 separate classifiers, used to detect gastrointestinal
- *Data extracted from records:*
    - ICD-9 diagnoses
    - Triage diagnoses (free text)
- *Case definitions (if applicable):*
    - 16 ICD-9 codes were included
    - Bayesian classifiers used to classify free-text triage diagnoses
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
    - Data were used to determine how accurate 3 separate classifying systems were at identifying acute gastrointestinal syndromes of public health interest
        - Compared to expert review of records
        - 1425 reports were randomly selected for review
    - Calculated indicators of accuracy for all 3 classifiers sets
- *Statistics/Results:*
    - 22/1425 cases (1.5%) fit case definition of acute GI of public health interest
    - Sensitivity ranged from 0.32 (ICD-9)-0.68 (Naïve Bayes)
    - Specificity ranged from 0.94 (Both Bayes)-0.99 (ICD-9)
    - ICD-9 classifier: 15 false negatives

- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Accuracy:
    - 98%: ICD-9
    - 94%: Naïve Bayes
    - 93%: Bigram Bayes
    - Positive predictive values were fairly low (0.10-0.37)

---

5. ***Title:***
Detection of Pediatric Respiratory and Gastrointestinal Outbreaks from Free-Text Chief Complaints.

Citation:
Ivanov O, Gesteland PH, Hogan W, Mundorff MB, Wagner MM. Detection of Pediatric Respiratory and Gastrointestinal Outbreaks from Free-Text Chief Complaints. *AMIA Annual Symposium Proceedings* 2003;2003:318-322.

Summary:
- *Data source:* **Emergency department** data gathered from a pediatric ED
  - **Inpatient** hospital admission data were also used for comparison of timeliness
    - Primary Children's Medical Center (PCMC) was data source: geographic area covered 80% of Utah's entire population
    - Data used from 1998-2001
- *Objectives of project:* To determine the efficacy of chief complaint data from a pediatric ED in ascertaining information about potential outbreaks
  - Respiratory outbreaks
  - Gastrointestinal outbreaks
- *Data extracted from records:*
  - Free-text chief complaint
  - Hospital admissions data
- *Case definitions (if applicable):*
  - Respiratory and gastrointestinal outbreaks were defined based on Complaint Coder (CoCo) Bayesian classification system (ED data)
  - Respiratory and gastrointestinal outbreaks were defined based on established set of ICD-9-CM codes (inpatient data)
    - Respiratory: pneumonia, influenza
    - GI: rotavirus, pediatric gastroenteritis
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
  - Data were used to determine the efficacy of chief complaint fields
  - Data were used to see timeliness of chief complaint availability in comparison to hospital admission data *gold standard (more timely detection of outbreaks)
  - Used Bayesian classification system (CoCo) to identify syndrome groupings
    - Daily counts for each of the two syndromes were calculated
  - Compared date of admission to hospital to date of ED chief complaint
  - Calculated cross-correlation of two time series (to ascertain lag time that maximized the correlation between ED and inpatient data)
  - Created a detection algorithm
    - Ran algorithm on "gold standard" inpatient data, again on ED

data, looked for lowest alert threshold that did not cause false positive alerts

- *Statistics/Results:*
  - Mean timeliness of outbreak detection
    - Respiratory: 10.3 days, 95% CI (-15.15, 35.5)
      - Not statistically significant, may be due to acquired nosocomial infections (since discharge data is based off ICD-9-CM codes, could have contracted respiratory illness in hospital even though initially admitted for a different reason)
    - GI: 29 days, 95% CI (4.23, 53.7)
  - Cross-correlation results:
    - 3 respiratory outbreaks detected
      - Timeliness: 7.4 days (mean), 95% CI (8.34-43.3)
    - 3 gastrointestinal outbreaks detected
      - Timeliness: 17.6 days (mean), 95% CI (3.4, 46.7)
  - ED visits came on average:
    - 10.3 days prior to hospital admissions (respiratory outbreaks)
    - 29 days (gastrointestinal outbreaks)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: chief complaint data provides important outbreak information several days prior to the availability of hospital inpatient data
  - Accuracy: Thresholds (number of cases) existed for both respiratory and GI that had 100% specificity and 100% sensitivity
    - Respiratory: 10-12 cases
    - GI: 5-9 cases

---

6. ***Title:***
   Self-reported fever and measured temperature in emergency department records used for syndromic surveillance.

   Citation:
   Kass-Hout TA, Buckeridge D, Brownstein J, Xu Z, McMurray P, Ishikawa CKT, Gunn J, Massoudi BL. Self-reported fever and measured temperature in emergency department records used for syndromic surveillance. *Journal of the American Medical Informatics Assocation* 2012; epub doi: 10.1136/amiajnl-2012-000847

   Summary:
   - *Data source:* **Emergency department** triage data gathered from Jan 2011-Oct 2011; retrieved from BioSense data feed, from a total of 665 US hospitals
     - Geographically diverse data (not specified)
   - *Objectives of project:* to measure frequency that measured temperature is recorded and compare measured temperature to self-reported fever
   - *Data extracted from ED records:* age, race, ethnicity, disposition (admitted or discharged), chief complaint, measured temperature
   - *Case definitions (if applicable):* Measured temperature was first assessed; if there was a measured temperature, chief complaint data was parsed to see if fever was listed
   - *Data use:* Information for health services/situational awareness
   - *Data analysis:*
     - Data was utilized to see if there is an association between measured

temperature in the ED and a self-reported fever listed in the chief complain section of the ED record
    - Fever is used in many syndromic surveillance case definitions—why it was studied here
    - Measured temperature was not usually recorded in ED records
    - Only approximately 13% of ED records received had a recorded temperature
- *Statistics*
    - 59.82% of patients with measured temperature also had self-reported fever
    - Only 36.94% of patients with self-reported fever had measured temperature
- *Discussion:* when only self-report was there, measured temperature was not usually elevated
    - Recording temperature more consistently in the ED could greatly aid syndromic surveillance efforts
- *Data quality:*
    - Timeliness: BioSense is an automated surveillance system so data was coming in near-real time

---

7. ***Title:***
Inpatient data: a new frontier in Veteran's Affairs biosurveillance and public health monitoring from the electronic health record.

Citation:
Lucero C, Schirmer P, Oda G, Holodniy, M. Inpatient data: a new frontier in Veteran's Affairs biosurveillance and public health monitoring from the electronic health record. *Emerging Health Threats Journal.* 2011;4:11063.

Summary:
- *Data source:* **Inpatient** data from Veterans Affairs ESSENCE system (all VA acute care hospitals)
- *Objectives of project:* To determine how VA ESSENCE inpatient data can be utilized for biosurveillance
- *Data extracted from records:*
    - ICD-9 codes
    - Demographic data
    - Admission/Discharge data
    - Diagnosis-related group
    - Bedsection
    - Procedure data
    - Surgery data
- *Case definitions (if applicable):*
    - N/A: utilized diagnosis code
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data were used to look for influenza and 1 hospital-acquired infection (HAI)→ C. difficile
    - Performed query on data looking for influenza diagnosis code
        - These results were compared to an optional CDC record of inpatient hospitalizations for influenza

    - Performed query on C. difficile diagnosis code
        - Calculated total monthly discharges for denominator rate
- *Statistics/Results:*
    - Average C. difficile rate was 11.1/1,000 discharges (national average of 78/100,000)
    - Influenza rates in inpatient data showed high correlation with CDC data (precise correlation not included in published abstract)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: gathered electronically, sent regularly (timeframe not specified)
    - Authors note that timeliness could be improved

---

8. ***Title:***
Adult asthma exacerbations and environmental triggers: a retrospective review of ED visits using an electronic medical record.

Citation:
May L, Carim M, Yadav K. Adult asthma exacerbations and environmental triggers: a retrospective review of ED visits using an electronic medical record. *American Journal of Emergency Medicine* 2011;29(9):1074-1082.

Summary:
- *Data source:* **Emergency department** records in the Washington, D.C. area
    - Total of 56,747 ED visits during study period
        - 554 visits were for asthma
        - 1,514 visits were for upper respiratory infection (URI)
    - Study period: June 1, 2005-May 30, 2006
- *Objectives of project:* To study the effects of various environmental allergens on the exacerbation and severity of asthma symptoms in adults (Acute Asthma Exacerbation—AAE)
- *Data extracted from records:*
    - Disposition diagnosis
    - Chief complaint
- *Case definitions (if applicable):*
    - Patients 18 years of age and older who were assigned one of the following ICD-9 billing codes (disposition diagnosis):
        - Asthma
        - Asthma exacerbation
    - Chief complaints (not further specified apart from these 2 symptoms):
        - Asthma
        - Wheezing
- *Data use:* Situational awareness/information for health services
- *Data analysis:*
    - Data were gathered to find rates in the ED of URI and asthma
    - Rates of allergens in the air were also collected for comparison (to monitor for exacerbations based on allergen levels)
    - Data were analyzed for their correlation with allergens and other environmental factors
    - Least squares regression model was created using backward stepwise modeling (using asthma visits and asthma admissions as the outcome variable of interest, environmental/allergen factors as independent

variables)
- *Statistics/Results:*
  - Asthma visits positively correlation with tree pollen amounts, average humidity: (R^2 = 0.631)
  - Asthma admissions negatively correlated with average temperature (adjusted for fine particulate matter): (R^2= 0.480)
  - AAE follow closely with tree pollen count trends
  - Asthma-related hospital admissions trend closely with colder weather (influenza season=possible reason)
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
  - Timeliness: data were analyzed retrospectively but could be available as soon as chief complaint information is filled in
  - Validity: included ED records from one ED only
  - Accuracy: 92% of patients with chief complaint containing "asthma" or "wheezing" were diagnosed with asthma in coded diagnosis

---

9. ***Title:***
Emergency Department Chief Complaint and Diagnosis Data to Detect Influenza-Like Illness with an Electronic Medical Record.

Citation:
May LS, Griffin BA, Bauers NM, Jain A, Mitchum M, Sikka N, Carim M, Stoto MA. Emergency Department Chief Complaint and Diagnosis Data to Detect Influenza-Like Illness with an Electronic Medical Record. *Western Journal of Emergency Medicine* 2010;11(1):1-9.

Summary:
- *Data source:* **Emergency department** records from urban ED with over 60,000 patients per year
  - South Atlantic region of United States (not specified further)
- *Objectives of project:* Determine if surveillance is more accurate at detecting influenza-like illness when it uses ICD-9 diagnosis codes in addition to chief complaint
- *Data extracted from records:*
  - Chief complaint data
  - ICD-9 diagnosis codes
- *Case definitions (if applicable):* Used case definitions close to CDC case definitions for the following syndromes:
  - Respiratory
  - Gastrointestinal
  - Viral illness
    - fever, headache, dry cough, sore throat, rhinorrhea, and myalgias
- *Data use:* Outbreak detection/situational awareness/information for health services
- *Data analysis:*
  - Data was gathered retrospectively from ED electronic medical records gathered between June 2005-May 2006
    - Some chief complaints were entered as free text, some entered in drop-down menu

    - 3 separate algorithms were applied to the data to find the onset date of the influenza season in the 2005-2005 season
- *Statistics/Results:*
    - 29% of discharge diagnoses differed from chief complaint
        - Disagreement differed by syndrome category (lowest disagreement for viral illness: 24%; highest disagreement for GI: 39%)
    - Diagnosis data was able to find an aberration in ILI data one month prior to chief complaint data
    - URI chief complaint frequency increased before URI diagnoses began increasing in frequency (may be more timely indicator)
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
    - Completion: reports were gathered from all but 5 of the 365 days during study period (4 of these days were not during flu season)
    - Accuracy: spikes in ILI mirror spikes in CDC national data for the region

---

10. ***Title:***
Peat Bog Wildfire Smoke Exposure in Rural North Carolina is Associated with Cardiopulmonary Emergency Department Visits Assessed through Syndromic Surveillance.

Citation:
Rappold AG, Stone SL, Cascio WE, Neas LM, Kilaru VJ, Carraway MS, Szykman JJ, Ising A, Cleve WE, Meredith JT, Vaughan-Batten H, Deyneka L, Devlin RB. Peat Bog Wildfire Smoke Exposure in Rural North Carolina is Associated with Cardiopulmonary Emergency Department Visits Assessed through Syndromic Surveillance. *Environmental Health Perspectives* 2011;119(10):1415-1420.

Summary:
- *Data source:* **Emergency Department** data from eastern North Carolina, pulled from NC DETECT syndromic surveillance program
- *Objectives of project:* To identify cardiopulmonary exacerbations following a peat bog wildfire, using an established syndromic surveillance system
- *Data extracted from records:*
    - Daily counts of ED visits
    - County of residence
    - Sex
    - Age
    - Date of admission
    - Discharge ICD-9-CM (looking for specific, pre-identified cardiac and respiratory codes)
- *Case definitions (if applicable):*
    - ICD-9-CM codes fitting noted respiratory and cardiovascular symptoms
        - Ex: asthma, heart failure, COPD
- *Data use:* Information for health services/situational awareness
- *Data analysis:*
    - Data were used to monitor for exacerbations of cardiac and respiratory illness following a wildfire in North Carolina
    - Diagnostic codes were used to identify cardiopulmonary “symptoms”
    - Data were pulled from onset of wildfire to first large rainfall that quelled the wildfire (6 week period from June 1-July 14, 2008)

    - Exposure rates to the smoke, gathered on a county level, were used for comparison
    - Normal existing background rates of cardiopulmonary illness were also noted
    - Calculated risk ratio for exposure to high levels of smoke for 3 days in a row (with 5 day lag period) compared to background rates
    - Calculated risk ratio for same 5 day lag period in reference counties that did not experience smoke from wildfire (or experienced smoke for up to 1 day)
- *Statistics/Results:*
    - Risk Ratio for respiratory illness in exposed counties=1.66, 95% CI (1.38-1.99)
        - RR for non-exposed counties=1.06, 95% CI (0.89-1.25)
    - Risk ratio for cardiac illness in exposed counties=1.37, 95% CI (1.01-1.85)
    - Overall risk ratio in counties exposed to smoke for cardiopulmonary-related events was statistically significant; RR=1.23, 95% CI (1.06,1.43)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: data were collected retrospectively
    - No data available on completeness
    - Accuracy: fairly accurate because utilizing ICD-9 diagnostic codes rather than less specific syndrome definitions

---

11. ***Title:***
Should we be worried? Investigation of signals generated by an electronic syndromic surveillance system—Westchester County, New York.

Citation:
Terry W, Ostrowsky B, Huang A. Should we be worried? Investigation of signals generated by an electronic syndromic surveillance system—Westchester County, New York. *Morbidity and Mortality Weekly Report* 2004;53 (Suppl):190-5.

Summary:
- *Data source:* **Emergency Department** records from Westchester County, New York
    - 7 total EDs (out of 13 in the county)
    - Data collected from end of Jaunary-end of October, 2003
- *Objectives of project:* To determine the importance and relevance of syndrome surveillance system signals and the time required by public health to respond to potential problems
- *Data extracted from records:*
    - Chief complaint (free text)
    - Hospital name
    - Age
    - Sex
    - Medical record number
    - Municipality
    - Patient zip code
    - Visit date
    - Discharge info (i.e., discharged from ED or admitted to inpatient)
- *Case definitions (if applicable):*

    - Used syndrome definitions developed by CDC, local health departments to identify a total of 8 different syndromes
- *Data use:* Information for health services
- *Data analysis:*
    - Data were used to see efficacy of system as well as to plan for public health response
    - Data were analyzed to look for statistically significant increases in number of visits in each of 8 syndrome categories
- *Statistics/Results:*
    - 59 statistically significant signals were detected
    - 32 of the signals were followed up by investigation by Westchester County Department of Health
    - No events of public health significance were detected
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Accuracy: after chart review, many automatic syndrome groupings were deemed to be incorrect
    - Completeness: 70% of total ED visits in the county were covered by system
    - Timeliness: data collected every 24 hours
        - On 8/277 days the data were not sent until the following day (a day late)

RETURN TO TABLE OF CONTENTS

Hospital and Ambulatory

This section, *Hospital and Ambulatory,* describes research and practice in using data from both hospital and ambulatory EHR records for syndromic surveillance purposes. These articles and abstracts highlight the potential of combing data streams from multiple clinical sources for improved syndromic surveillance.

12. ***Title:***
Local influenza-like illness surveillance at a university health system during the 2009 H1N1 influenza pandemic.

Citation:
Baker AW, Enfield K, Mehring B, Turner JC, Sifri CD. Local influenza-like illness surveillance at a university health system during the 2009 H1N1 influenza pandemic. *American Journal of Infection Control.* 2012 (available online ahead of print version) http://dx.doi.org/10.1016/j.ajic.2011.12.009.

Summary:
- *Data source:* 8 sentinel surveillance sites (**ambulatory** and **emergency department**) from University of Virginia Health System
    - 2009-2010 influenza season
    - 2.5 million people are seen by health system
- *Objectives of project:* To assess the efficacy of a university health care system at detecting influenza in the region
- *Data extracted from records:*
    - Discharge diagnosis (ICD-9 code) (ED)
    - ICD-9 diagnosis code (student health center)
    - Classification by nurse as ILI-related visit (ambulatory—not student health center)
- *Case definitions (if applicable):*
    - 6 of 8 sites:
        - fever (temperature greater than 38 degrees celsius)
        - cough, sore throat, nasal congestion OR body aches
        - exclude: other known cause of these symptoms
    - Student health clinic:
        - ICD-9 codes for influenza with pneumonia, influenza with other respiratory manifestations
    - ED
        - Discharge diagnosis with terms: flu, influenza, viral illness or viral syndrome
- *Data use:*
    - Outbreak detection/situational awareness
- *Data analysis:*
    - ILI rates through university health system were compared to region and state-wide ILI data
    - Compared ILI rates through university health system to Health and Human Services rates from the region and nationally
    - Compared 6 sites using symptom case definition to 2 sites using diagnostic codes
    - Local ILI network cases were compared to influenza test results

- *Statistics/Results:*
  - R^2 = 0.88 for agreement between symptom-based case definition locations and diagnostic code-based definition locations
  - R^2 for agreement between:
    - Local ILI network and Virginia regional information (0.96)
    - Local ILI network and state of Virginia (0.93)
    - Local ILI network and Health and Human Services Regional data (0.94)
  - Flu cases confirmed through laboratory testing followed a similar pattern to syndromic data (local network only)
    - Peak of laboratory results data was 2 weeks prior to peak of syndromic data
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: weekly data was sent Monday (gathered for prior week of Monday-Sunday)→ sent within 24 hours of end of week
  - Timeliness: during the high point of influenza season data was gathered daily and reports were put out twice each week
  - Accuracy: did not include fever in syndrome definition to increase sensitivity

---

13. ***Title:***
Automated surveillance of Clostridium difficile infections using BioSense.

Citation:
Benoit SR, McDonald LC, English R, Tokars JI. Automated surveillance of Clostridium difficile infections using BioSense. *Infection Control and Hospital Epidemiology* 2011. 32(1):26-33

Summary:
- *Data source:* **Inpatient, Emergency Department** and **Ambulatory** care laboratory data
  - BioSense data streams from 44 total hospitals
  - Data pulled from January 1, 2007-June 30, 2008
  - Geographic region not specified
- *Objectives of project:* To assess the new surveillance definition for *C. difficile* infection (CDI) and assess the ability of BioSense data to accurately determine CDI rates
- *Data extracted from records:*
  - Results from laboratory testing from *C. difficile* positive-toxin assays
  - Age
  - Sex
  - Race
  - Ethnicity
  - ICD-9-CM code
  - Setting
  - Category (Hospital Onset, Community-Onset Hospital-Associated, Community-Onset Non-Hospital-Associated)
- *Case definitions (if applicable):*
  - Hospital-onset CDI
    - *C. difficile* toxin-positive result 3 or more days after hospital admission

    - Community-onset CDI
        - *C. difficile* toxin-positive result less than 3 days after hospital admission (or in the absence of a hospital admission)
        - Hospital-Associated: overnight admission occurred in previous 30 day period
        - Non-Hospital Associated: no overnight admission occurred in previous 30 day period
- *Data use:* Outbreak detection
- *Data analysis:*
    - Total number of patients with *C. difficile* toxin-positive laboratory result was calculated
        - Percentage of cases that were hospital-onset and community-onset was also calculated
    - Results compared to known rates of CDI
- *Statistics/Results:*
    - 4,585 unique patients with CDI were identified
        - 282 (6.2%) were identified using coded elements, the rest were identified using free-text laboratory results
    - 47% of cases were hospital-onset
    - 69.2% of community-onset cases were non-hospital-associated
    - 10.5 cases of CDI/10,000 patient-days
        - rate of hospital-onset CDI: 7.8 cases/10,000 patient-days
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Accuracy: 76.2% of inpatients with positive laboratory result had ICD-9-CM code in record designating a CDI
        - Rates dropped by setting: 25.3% of outpatients, 28% of ED patients
    - Timeliness: study was performed retrospectively
    - Validity: other studies have found similar rates of CDI
        - Hospital-onset CDI is a reportable disease in Ohio
            - Rates 6.4-7.9 cases/10,000 patient-days

---

14. ***Title:***
Classification of Emergency Department Chief Complaints Into 7 Syndromes: A Retrospective Analysis of 527,228 Patients.

Citation:
Chapman WW, Dowling JN, Wagner MM. Classification of Emergency Department Chief Complaints Into 7 Syndromes: A Retrospective Analysis of 527,228 Patients. *Annals of Emergency Medicine.* 2005; 46(5): 445-455.

Summary:
- *Data source:*
    - 527,228 triage chief complaint records from University of Pittsburgh Medical Center **Emergency Department** (1990-2003)
- *Objectives of project:* To verify the accuracy of a chief complaint Bayesian classification system (CoCo) by comparing it with ICD-9 discharge diagnoses
- *Data extracted from records:*
    - Free text triage chief complaint
    - ICD-9 discharge diagnosis (as gold standard for comparison)
- *Case definitions (if applicable):* CoCo case definitions exist for 7 syndrome

categories, over 10,000 chief complaint keywords are coded to fit in with the following syndrome categories based on probability (i.e., a chief complaint is put in the most likely syndrome category):
    - Respiratory
    - Gastrointestinal
    - Rash
    - Botulinic
    - Hemorrhagic
    - Neurologic
    - Constitutional
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
    - Data were analyzed to determine accuracy of syndrome groupings
    - Data were analyzed to determine generalizability to other geographic regions
- *Statistics/Results:*
    - 85,569 patients were classified into a syndrome category
    - 26% of patient records were incorrectly classified by CoCo system
        - 31% of false negatives incorrectly classified
        - 22% of false positives incorrectly classified
    - 14% of overall misclassification errors were due to CoCo system errors
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Accuracy: 92%-99% depending on syndrome grouping
    - Sensitivity: 30%-75%
    - Positive predictive value: 12%-44%
    - External validity: a sample test was performed using ED information from Utah, suggests that there is little to no geographic variability (high external validity)
    - Timeliness: Chief complaint fields are readily available though may not contain all relevant information)

---

15. ***Title:***
Fever detection from free-text clinical records for biosurveillance.

Citation:
Chapman WW, Dowling JN, Wagner MM. Fever detection from free-text clinical records for biosurveillance. *Journal of Biomedical Informatics*. 2004;37(2):120-127.

Summary:
- *Data source:* **Emergency department** medical records
    - University of Pittsburgh Medical Center
- *Objectives of project:* To determine which of 3 free text parsing systems were best suited to identifying febrile illness syndrome
- *Data extracted from records:*
    - Free text chief complaint field
- *Case definitions (if applicable):*
    - Measured temperature of 38.0 degrees Celsius or higher OR
    - Description of recent fever or chills
- *Data use:* Information for health services/outbreak detection
- *Data analysis:* 3 separate algorithms were evaluated
    - Keyword HP

        - Searches for negation to eliminate false positives
        - Searches for 2 elements of case definition (using fever keywords)
    - Keyword CC
        - Searched for fever keywords and term temp*
    - Keyword CoCo
        - Searched for fever
        - Searched for constitutional symptoms (fever is classified by CoCo as a constitutional symptom): constitutional symptoms other than fever may also be indicative of fever
    - Randomly assigned test set of 213 patients was collected
    - 50% of data were pulled from patients with the ICD-9 code for fever (780.6), 50% of data were pulled from patients without ICD-9 code for fever
- *Statistics/Results:*
    - Sensitivity ranged from 57% (CoCo) to 98% (HP)
    - Specificity ranged from 89% (HP) to 100% (CC)
    - Prevalence of fever=51%
    - Keyword HP was statistically significantly better than CC or CoCo (p-value<0.05)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Accuracy: Keyword HP most sensitive (Sens=98%)
    - Accuracy: Keyword CC most specific (Spec=100%)
    - Completeness: only 4% (n=9) of records did not contain information pertaining to febrile status
        - 88% of patients with fever had measured temperature recorded in clinical record
    - External validity: should be applicable to areas outside of biosurveillance (study authors)
    - External validity: only contained information from a single health care system (negative)

---

16. ***Title:***
Establishing an emergency department syndromic surveillance system to support the London 2012 Olympics and Paralympic Games.

Citation:
Elliot AJ, Hughes HE, Hughes TC, Locker TE, Shannon T, Heyworth J, Wapling A, Catchpole M, Ibbotson S, McCloskey B, Smith GE. Establishing an emergency department syndromic surveillance system to support the London 2012 Olympics and Paralympic Games. *Emergency Medicine Journal* 2012; doi:10.1136/emermed-2011-200684

Summary:
- *Data source:*
    - Extracted from daily feeds from 6 England **Emergency Departments**
- *Objectives of project:* To establish if Emergency Department Sentinel Syndromic Surveillance System (EDSSS) is an effective and feasible way to monitor public health during the Olympics
- *Data extracted from records:* age, gender, partial postcode, triage coding, diagnosis codes
- *Case definitions (if applicable):* N/A—only looked at overall number of

admissions
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:* Analyzed ED attendance by hour of day and day of week; winter season saw increase in incidence of respiratory and ARI infections in ED
- *Statistics:*
  - Many related to data quality
  - Average of 527 mean daily attendance (range: 458-614) in beginning (prior to implementation of all 6 ED sites)
  - Highest day time attendance on Mondays
  - Highest time of attendance (7 am-10 am)
- *Data quality:*
  - Completion rates were fairly high (recording of triage presentation ranged from 0-100%); poorest completion in diagnostic code field, <66% overall (range: 17-89%)
  - Patient sex: 100% mean completion rate
- Passive requirement (EHR system was already set up, data is now automatically gathered and sent to EDSSS)

---

17. ***Title:***
Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness.

Citation:
Espino JU, Wagner MM. Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. *Proceedings AMIA Symposium* 2001:164-8.

Summary:
- *Data source:* **Emergency Department** data from University of Pittsburgh Medical Center
- *Objectives of project:* To compare the accuracy of ICD-9 coded diagnoses and chief complaint to a manual classification of illness
- *Data extracted from records:*
  - ICD-9 coded chief complaint
  - ICD-9 coded diagnosis code
- *Case definitions (if applicable):* Acute Respiratory Illness (ARI) was defined as anyone showing symptoms for less than 5 days with any of the following symptoms:
  - cough, shortness of breath, sputum production, abnormal pulmonary examination, or radiological evidence of pneumonia
  - Cases that could be explained by a non-ARI event were excluded from the analysis
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
  - Any code that the coder believed would be given to someone presenting with ARI >/= 5% of the time was defined as indicative of ARI
  - Any patient presenting with one of the coded ICD-9 codes (in diagnosis field or chief complaint field) was categorized as having ARI
  - A sample of 800 cases was selected for analysis
  - Only most recent patient visit for each patient was included (de-duplication occurred)
  - 2 sets of physicians reviewed medical records and flagged them as

"positive" if there was evidence of ARI
    - physicians did not know patient ICD-9 code at time of record review
    - Any disagreement between the 2 phases of physician review was settled through a meeting and mutual agreement
- *Statistics/Results:*
    - Pearson's correlation for interrate reliability ranged from 0.759 to 0.941
    - Sensitivity: 44%
    - Specificity: 97%
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: Mean lag time between entering of ICD-9 code and availability was 7.5 hours (range: up to 80.6 hours)
    - Validity: Interrater reliability correlations were high (>0.75)
    - External validity not known
    - Accuracy: Sensitivity was relatively low, specificity extremely high (97%)

---

18. ***Title:***
Automated Syndromic Surveillance for the 2002 Winter Olympics.

Citation:
Gesteland PH, Gardner RM, Tsui F-C, Espino JU, Rolfs RT, James BC, Chapman WW, Moore AW, Wagner MM. Automated Syndromic Surveillance for the 2002 Winter Olympics. *Journal of the American Medical Informatics Association*. 2003;10(6):547-554.

Summary:
- *Data source:* **Acute care** and **emergency department** stream received in real-time, Feb 8-March 31, 2002 through 2 healthcare systems: University of Utah Health Sciences Center and Intermountain Health Care (IHC)
    - Geographic area is home to 2 million people
    - Univ of Utah: covers 10% of population
    - IHC: covers 60% of population
    - 19 urgent care centers, 9 EDs provided data
- *Objectives of project:* To implement and test a syndromic surveillance system, with an aim of investigating any events of public health significance that may arise
- *Data extracted from records:*
    - Chief complaint (free text)
    - Demographic data
    - Facility ID
    - Unique record ID
    - Gender
    - Age (years)
    - Visit date/time
    - Patient zip code
- *Case definitions (if applicable):*
    - Used established Real-time Outbreak and Disease Surveillance (RODS) syndrome definitions for 7 syndromes
- *Data use:* Outbreak detection
- *Data analysis:*
    - Regressive Least Square (RLS) adaptive filters were calculated to provide expected counts

        - Emphasizes recent history/frequency rather than distant past (better for use during large scale event bringing many new people to the area)
        - Signal occurred when difference between expected and actual RLS was statistically significant at a 95% confidence level ($p<0.05$)
- *Statistics/Results:*
    - 861 average ED visits per day
    - 1,386 average acute care visits per day
    - 2 signals were detected by the algorithm in place
        - both were determined to be false alarms
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: data received in real time
    - Timeliness: analysis occurred every 4 hours
    - Completeness: records from Univ of Utah and IHC covered 70% of all acute care and ED visits in the geographic area
    - Validity: used established HL-7 message format for high generalizability

---

19. ***Title:***
Gastrointestinal disease outbreak detection using multiple data streams from electronic medical records.

Citation:
Greene SK, Huang J, Abrams AM, Reed M, Platt R, Kulldorff M. Gastrointestinal disease outbreak detection using multiple data streams from electronic medical records. *Emerging Health Threats Journal*. 2011;4:s25.

Summary:
- *Data source:* 22 different **ambulatory, emergency department and inpatient** EMR data streams collected from Kaiser Permanente Northern California system (encompasses data from 3.3 million people)
- *Objectives of project:* To determine which electronic health information feeds (or combinations of feeds) are best suited to detecting gastrointestinal illness outbreaks AND to focus in on case definitions for GI syndrome
- *Data extracted from records:* 4 different types of data streams provided different types of information
    - Syndromic definitions (n=14)
    - Prescription drug dispensings (n=1)
    - Microbiology test ORDERS (n=1)
    - Microbiology test RESULTS (n=6)
    - All of the streams collected frequency data by residential zip code
    - Data types included:
        - free text
        - ICD-9 codes
        - test results
        - pharmacy prescription details (searching for antibiotics)
- *Case definitions (if applicable):* Case definitions were adapted from established CDC/Dept of Defense definitions
    - Upper GI=included vomiting
    - Lower GI=included diarrhea, gastroenteritis
- *Data use:* Outbreak detection

- *Data analysis:*
  - Data was compiled into 22 separate streams which were assessed separately and in combination
  - Patients with more than one encounter in 14 days were deduplicated and counted only once
  - Single stream analyses were performed on each of the 22 streams
  - Multivariate analyses were performed on 5 data streams (3 microbiology streams and 2 syndromic streams)
  - A list of known GI outbreaks in the area was created
- *Statistics/Results:*
  - 24 potential outbreaks were identified (using data from 16 non-microbiological streams)
    - 3 most effective streams (identified >5 outbreaks)
      - Lower GI: ED setting
      - Upper GI: ED setting
      - Lower GI: Ambulatory Care setting
  - 6 Pathogen-specific streams
    - were able to identify 5 outbreaks including one probably common-source outbreak (was not detected using other streams)
  - Multivariate analysis:
    - Was more timely than single-stream sources in at least one instance
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: data were analyzed in 24 hour increments to simulate a daily feed
  - Specificity: Microbiology tests are very specific
  - Validity: Creates a lot of noise that may not necessarily indicate a true outbreak BUT could create a focal point for local health departments to study

---

20. ***Title:***
Timely detection of localized excess influenza activity in Northern California across patient care, prescription, and laboratory data.

Citation:
Greene SK, Kulldorff M, Huang J, Brand RJ, Kleinman KP, Hsu J, Platt R. Timely detection of localized excess influenza activity in Northern California across patient care, prescription, and laboratory data. *Statistical Medicine*. 2001;30:549-559.

Summary:

- *Data source:* 10 Kaiser Permanente Northern California data streams gathered from:
    - **Ambulatory Care** records (1 stream with fever as a criterion, another without fever)
    - Prescription records (searching for antiviral medication dispensings)
    - **ED records** (1 stream with fever as a criterion, another without fever)
    - **Inpatient** records
    - Lab Orders
    - Lab Results
- *Objectives of project:* To compare various data streams in order to determine which are best at detecting ILI
- *Data extracted from records:*
    - ICD-9 diagnosis codes
    - Zip code
- *Case definitions (if applicable):* Definition used for ILI was the National Bioterrorism Syndromic Surveillance Demonstration Project definition (at least one of the following):
    - Viral infection
    - acute pharyngitis
    - acute laryngitis and tracheitis
    - acute bronchitis and bronchiolitis
    - other upper respiratory disease
    - pneumonia
    - influenza
    - throat pain
    - cough
    - *Fever was included with some data streams
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
    - Data were only used if there had been no encounter in the previous 42 days
    - Study simulated a prospective study by utilizing a space-time permutation scan statistic—allowed study authors to analyze data as if it were coming in near-real time data streams
    - Data were also examined retrospectively to detect ILI outbreaks
    - Data streams were analyzed searching for recurrence intervals (RIs) and signals were considered significant if the RI>365 days
    - Data streams showing RI signals between 100 days and 365 days were considered weaker but still potentially useful data sources
- *Statistics/Results:*
    - ILI Event #1
        - Signals in lab orders and ambulatory care stream (without fever)
    - ILI Event #2

        - Signals in prescription records, lab orders, ambulatory care (without fever)
    - ILI Event #3
        - 3 very strong signals: ambulatory care (without fever), ambulatory care (with fever), lab orders
    - ILI Event #4
        - Lab orders, lab results, ambulatory care (without fever), emergency department (without fever)
    - Ambulatory care without fever criterion detected all 4 ILI events of significance AND was the most timely method of detecting the event 2 of 4 times
        - Lab results, prescription records, and lab tests also provided very timely results
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Accuracy: no gold standard to compare to
    - Timeliness: timeliness was decreased when fever was added as a necessary criterion to ambulatory and ED feeds

---

21. ***Title:***
Disease outbreak detection system using syndromic data in the greater Washington DC area.

Citation:
Lewis MD, Pavlin JA, Mansfield JL, O'Brien S, Boomsma LG, Elbert Y, Kelley PW. Disease outbreak detection system using syndromic data in the greater Washington DC area. *American Journal of Preventive Medicine* 2002;23(3):180-186.

Summary:
- *Data source:* Department of Defense ESSENCE medical information systems, pulling records from **emergency departments** and **primary care clinics**
    - 99 total EDs and primary care clinics
    - Data received from December 1999-January 2002
    - Geographically diverse data (not specified)
- *Objectives of project:* To determine the efficacy of ED and ambulatory care data in identifying infectious disease outbreaks in a more timely fashion than other surveillance methods
- *Data extracted from records:*
    - ICD-9-CM codes (diagnostic codes)
    - Patient disposition
    - Procedures
    - Patient zip
    - "Other data'
- *Case definitions (if applicable):* ICD-9-CM codes were grouped into 7 total syndromes
    - Each syndrome is made up of 3 ICD-9-CM codes (most common examples):
        - GI includes: gastroenteritis, enteritis, nausea with vomiting
        - Respiratory includes: bronchitis, infection, pharyngitis
        - Fever includes: septicemia, septicemia (not otherwise specified), fever
- *Data use:* Outbreak detection

- *Data analysis:*
  - Data are used to develop historical counts
  - Monitor for outbreaks/alerts
  - Calculate syndrome frequency
  - Correlations were calculated (between syndromic data and CDC data)
  - Predicted and actual counts were calculated
  - 95% confidence interval upper boundaries were calculated (to form alerts)
  - Geomapping by patient zip
- *Statistics/Results:*
  - Correlation between respiratory syndrome data and CDC data for southeast US during same time period=0.704 ($p<0.001$)
  - Detected 3 GI outbreaks in separate cities (January 2002)
    - Determined to be true rotavirus outbreak
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: receive data daily, data analyzed within 1-3 days of patient encounter
  - Timeliness: noted flu numbers increasing 6 weeks earlier than the previous year→ syndromic system was able to track this pattern before it was verified by CDC data
  - Validity: accurately detected GI syndrome in real-time during 3 separate rotavirus outbreaks

---

22. ***Title:***
Enhanced health event detection and influenza surveillance using a joint Veteran's Affairs and Department of Defense biosurveillance application.

Citation:
Lucero C, Oda G, Cox K, Maldonado F, Lombardo J, Wojcik R, Holodniy. Enhanced health event detection and influenza surveillance using a joint Veteran's Affairs and Department of Defense biosurveillance application. *BMC Medical Informatics and Decision Making*. 2011;11:56.

Summary:
- *Data source:*
  - Combined data from 2 different source populations: Department of Defense and Veterans Affairs
  - Data gathered from Chicago, Illinois
  - **Ambulatory, emergency department, limited inpatient records**
- *Objectives of project:* To evaluate how well a biosurveillance system that combined the 2 data sources performed
- *Data extracted from records:*
  - Total number of visits per month
  - Patient age
  - Patient gender
  - Syndrome group
  - ICD-9 diagnosis code
- *Case definitions (if applicable):*
  - Case definitions are not presented in this article but the syndrome categories are:
    - Botulism-like

        - Febrile disease
        - Fever
        - GI
        - Hemorrhagic illness
        - ILI
        - Miscellaneous/Asthma
        - Neurological
        - Rash
        - Respiratory
        - Shock/Coma
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data elements were combined and analyzed focusing on syndrome group and ICD-9 diagnosis code
    - Major outbreaks in Chicago area during study time (June 2006-March 2009) were identified for comparison to see if systems detected outbreak events
    - Duplicates were identified and removed
    - Descriptive statistics and univariate analyses were performed
    - Regression/Exponentially Weighted Moving Average was calculated for comparing data to major identified outbreaks
    - Weekly percentage of visits that related to ILI were compared to CDC numbers for the same weeks
- *Statistics*
    - High level alerts were able to identify many of the outbreaks that had occurred during the study time period
    - Combined system did not appear to provide more accurate information than each of the systems had provided on their own
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: retrospective study
    - Completion: article states that "all records" were pulled, indicating a completion rate close to if not 100%

---

23. ***Title:***
A comparison of ambulatory care and emergency department encounters as data sources for detection of clusters of lower gastrointestinal illness.

Citation:
Yih WK, Abrams A, Hsu J, Kleinman K, Kulldorff M, Platt R. A comparison of ambulatory care and emergency department encounters as data sources for detection of clusters of lower gastrointestinal illness. *Advances in Disease Surveillance* 2006;1:75.

Summary:
- *Data source:* **Ambulatory care** (AC) and **Emergency Department** (ED) data from Kaiser Permanente Northern California (KPNC) and Harvard Pilgrim Health Care-Harvard Vanguard Medical Associates (HPHC-HVMA) from 2003
    - KPNC: 3 million
    - HPHC-HVMA: 140,000
- *Objectives of project:* To compare the detection abilities of ambulatory care and emergency department data streams to detect lower gastrointestinal illness
- *Data extracted from records:*

    - Diagnostic code
    - Zip codes
- *Case definitions (if applicable):*
    - CDC syndromic definition for lower GI
    - Repeat visits within 6 weeks, to same setting, for same symptoms, were removed from data set
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data from each stream (ambulatory and emergency department) were compared separately
    - Also looked at detection capacity of the two data streams together
    - Recurrence intervals were calculated (considered significant if RI>30 days)
    - Frequency of signals from the different streams were calculated (4 streams in all: AC and ED data from each of 2 sites)
- *Statistics/Results:*
    - KPNC
        - 13 signals: ambulatory
        - 4 signals: ED
        - 10 signals: ambulatory + ED
    - HPHC-HVMA
        - 3 signals: ambulatory
        - 7 signals: ED
        - 14 signals: ambulatory + ED
    - 30% of signals required both ED and AC data to be detected
    - No overlapping signals between AC and ED data
        - Indicates that each data stream has unique usefulness in syndromic surveillance
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: data collected retrospectively
    - Accuracy: utilized validated CDC definition for lower GI syndrome

---

24. ***Title:***
Evaluating Syndromic surveillance systems at institutions of higher education (IHEs): A retrospective analysis of the 2009 H1N1 influenza pandemic at two universities.

Citation
Zhang Y, May L, Stoto MA. Evaluating Syndromic surveillance systems at institutions of higher education (IHEs): A retrospective analysis of the 2009 H1N1 influenza pandemic at two universities. *BMC Public Health*. 2011;11:591.

Summary:
- *Data source:* Data from 2 universities in Washington, D.C. including reports of influenza-like illness, employee absenteeism rates and hygiene supply distribution records
(included **ambulatory** and **emergency department** data)
    - Fall 2009
- *Objectives of project:* To assess the effect of pH1N1 on university campuses and assess the ability of these ambulatory care monitoring systems to detect influenza outbreaks

- *Data extracted from records:*
  - Chief complaint (from ED records)
  - Discharge diagnosis (from ED records)
  - Symptom information (from ambulatory care centers)
- *Case definitions (if applicable):*
  - Case definition used for telephone call center, ambulatory care centers:
    - Fever of 100 degrees farenheit or greater AND
    - Cough and/or sore throat
    - Exclude other known causes of symptoms
  - ED visits case definition:
    - Discharge diagnosis of flu-related diagnosis
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:*
  - Used data gathered from university sources and compared to CDC ILI rates, American College Health Association (ACHA) attack rates, Google Flu trends
  - Data were analyzed using a triangulation technique to attempt to determine what results were as a result of biases and which indicated true trends in influenza-like illness
  - Data were compared to other data sources to note trends
- *Statistics/Results:*
  - Peaks from University A did not match CDC, ACHA data or Google Trends data (showed an additional peak)
  - University B: first peak showed ambulatory care visits, 2nd peak showed ED visits
    - Ambulatory care data may be useful for predicting severity (since the second peak was more severe)
  - Student health services data from both universities are comparable
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Accuracy: difficult to know if influenza-like illness (or even verified influenza) is pH1N1, the specific strain being monitored
  - Accuracy: gold standard was developed through triangulation technique
  - Validity: students were asked to report when they became ill with influenza-like symptoms (may have an effect on validity)

RETURN TO TABLE OF CONTENTS

Ambulatory

This section, *Ambulatory,* describes research and practice in using data from ambulatory EHR records for syndromic surveillance purposes. For the purposes of this annotated bibliography, both in-person and telephone clinical encounters were considered to be applicable.

25. ***Title:***
General Practice Out-of-Hours Service in Ireland Provides a New Source of Syndromic Surveillance Data on Influenza.

Citation:
Brabazon ED, Carton MW, Murray C, Hederman L, Bedford D. General Practice Out-of-Hours Service in Ireland Provides a New Source of Syndromic Surveillance Data on Influenza. *Eurosurveillance*. 2010;15(31):1-7.

Summary:
- *Data source:* **General practitioner (ambulatory)** Out-of-Hours (OOH) services (Ireland)
    - Telephone-based; open hours when GP offices are not (including weekends and holidays)
- *Objectives of project:* To determine the occurrence of ILI according to OOH EHR records and compare that to national data containing established ILI rates (that are gathered utilizing sentinel GP systems)
- *Data extracted from records:*
    - Keywords were extracted from the 'patient's reported condition' field of the EHR.
    - Data with the following keywords→ 'fever' or 'high temperature' and **2 or more of:** 'headache' 'sore throat' 'cough' 'aches and pains' were determined to be ILI
    - Extracted from aggregated data set containing all calls from one week periods (Monday through Sunday)
- *Case definitions (if applicable):*
    - fever/elevated temperature and 2 or more of:
        - headache, sore throat, cough, or aches and pains
- *Data use:* Outbreak detection
- *Data analysis:*
    - Total number of calls were compared to national ILI system (no correlation)
    - Subset of data was coded to determine which definition of ILI would provide most accuracy. Each symptom of ILI was coded as either 0 (no symptom) or 1 (symptom present) and the string of binary codes was compared to 3 separate ILI definitions: Ireland's ILI definition, CDC's ILI definition, and a self-reported illness definition. Sensitivity and specificity for each of the three ILI definitions were compared.
        - **Self-reported ILI** had the least noise and extracted the most records so that was used as the comparative measure for the study
- *Statistics*
    - Over 539,000 calls were made to 2 OOH services during 6 influenza seasons (end 2003-beginning of 2009)
    - Correlations were all statistically significant (highest Spearman's rank correlation was 0.909 in 2003-2004 influenza season): Correlations

compared self-reported ILI retrieved from OOH services to national ILI rates gathered separately
- *Data quality*
  - Quality checks occurred when the three different ILI definitions were tested for specificity, sensitivity, and positive predictive value
    - Self-reported ILI was utilized after it recovered less noise than the Irish ILI definition and recovered more records than the CDC ILI definition

---

26. ***Title:***
Coverage and Timeliness of Combined Military and Veteran Surveillance Systems.

Citation:
Burkom HS, Elbert Y, Winston C, Pavlin J, Lucero-Obusan C, Holodniy M. Coverage and Timeliness of Combined Military and Veteran Surveillance Systems. *Online Journal of Public Health Informatics.* 2013;5(1):29.

Summary:
- *Data source:*
  - Outpatient record information from Department of Veterans Affairs (VA) and Department of Defense (DoD) surveillance systems from 2007-2010
- *Objectives of project:* To determine how best to combine outpatient data from the VA with data from the DoD; looked at coverage and timeliness
- *Data extracted from records:*
  - ICD-9 codes
  - Patient age
  - Geographical information on statistical area
    - Megapolitan (> 1 million)
    - Metropolitan (500,000-1 million)
    - Micropolitan (10,000-50,000)
- *Case definitions (if applicable):* N/A
- *Data use:* increase efficacy and awareness of outputs from outpatient data streams
- *Data analysis:*
  - Looked at total coverage, areas with dual coverage (DoD AND VA)
  - Compared symptom-based ILI rates
- *Statistics/Results:*
  - VA coverage is overall greater
  - In areas with dual coverage, DoD visit volume is generally greater
  - >85% of VA patients are over the age of 45; only 22% of DoD patients are over the age of 45
  - DoD uses symptom-based ILI codes more frequently than VA
  - Detection algorithms—generally better in DoD data (57% of facilities-seasonal influenza; 77% of facilities-pandemic influenza)
  - Detection algorithms are better in VA data for 37% of facilities-seasonal influenza; 14% of facilities-pandemic influenza
  - DoD data was better than VA data at detecting H1N1 and H3N2
  - Combined data streams increases efficacy at detecting influenza
-

- *Data quality:*
  - Timeliness: can be improved in 92% of areas if data from both types of facilities was used

---

27. ***Title:***
Validation of a syndromic surveillance system using a general practitioner house calls network, Bordeaux, France.

Citation:
Flamand C, Larrieu S, Couvy F, Jouves B, Josseran L, Filleul L. Validation of a syndromic surveillance system using a general practitioner house calls network, Bordeaux, France. *Eurosurveillance*. 2008;13(25).

Summary:
- *Data source:* Syndromic Surveillance system set up in France to retrieve information from **general practitioner** house calls in the area **(ambulatory)**
  - Data collected form 2005-2006; total of 303,396 visits occurred
  - Data collected from 60 local organizations that make house calls 24 hours a day, 7 days a week
- *Objectives of project:* To increase surveillance in the region by including ambulatory records and allow for tracking of illness that may not be serious enough for an ED visit or hospitalization
- *Data extracted from records:*
  - Date of visit
  - Postal code
  - Age
  - Sex
  - Health complaints
  - Diagnosis
- *Case definitions (if applicable):*
  - Syndrome definitions established by ICPC-2 Codes (International Classification of Primary Care)
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:*
  - Data analyzed for total volume
  - Data analyzed for specific syndrome groups (16 groups in all)
- *Statistics:*
  - Correlation between records retrieved from Bordeaux region only and entire Aquitaine region were compared (also utilized slightly different case definitions for ILI): Correlation between two systems was 0.92
  - Alert thresholds for heat-related illness showed significant correlation with increase in number of GP visits for heat-related illness: Correlation: 0.72 (p-value < 0.0001)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Completion rate: 85%
  - Timeliness: Data sent automatically every 24 hours
  - Accuracy: used established international syndrome definitions (though no comparison made to test results, etc.)
  - Validity: diagnosis data have been shown to be a better indicator of syndrome for syndromic surveillance than chief complaint data, and this system utilizes both for improved accuracy

28. ***Title:***
Comparison of Office Visit and Nurse Advice Hotline Data for Syndromic Surveillance—Baltimore-Washington, D.C., Metropolitan Area, 2002.

Citation:
Henry JV, Magruder S, Snyder M. Comparison of Office Visit and Nurse Advice Hotline Data for Syndromic Surveillance—Baltimore-Washington, D.C., Metropolitan Area, 2002. *Morbidity and Mortality Weekly Report (MMWR).* 2004; 53(Suppl):112-116.

Summary:
- *Data source:* Kaiser Permanente of the Mid-Atlantic States (KPMAS) nurse advice hotline data **(ambulatory)**
    - Calls from 2002 were considered
    - 500,000 members (including 30 outpatient centers)
    - 1,497,686 calls were made to hotline in 2002
- *Objectives of project:* To determine how well calls to the nurse advice hotline predicts the diagnosis of a patient who is later diagnosed in a physician in-person encounter
- *Data extracted from records:*
    - In-person encounter date
    - Appointment schedule date
    - Telephone encounter date/time
    - Patient ID
    - ICD-9 codes (for in person encounter only)
    - Advice guidelines selected by nurse; groups symptoms into a syndrome category (for telephone encounter only)
- *Case definitions (if applicable):* ESSENCE II syndrome definitions, utilizing ICD-9 diagnostic codes
- *Data use:* Information for health services/situational awareness/outbreak detection
- *Data analysis:*
    - Determine frequency of calls to KPMAS nurse hotline
    - Determine frequency of visits to KPMAS (in-person)
    - Determine correlation between syndrome groupings assigned over the phone and syndrome groupings assigned in-person
    - Data were analyzed for:
        - Frequency
        - Agreement between diagnosis syndrome categories
    - Telephone encounters that were in agreement with in-person diagnosis were analyzed to calculate average lag time between the telephone encounter and the in-person visit
- *Statistics/Results:*
    - Respiratory and GI produced greatest number of telephone and in-person encounters (GI: over 70,000 calls, over 26,000 visits; respiratory: over 240,000 calls, over 200,000 visits)
    - GI and respiratory syndromes had the highest number of syndrome matches between telephone and in-person encounter
        - Mean hours between phone call and visit for both syndromes was approximately 12 hours
    - Respiratory syndromes showed the highest positive predictive value

levels (63.8%) and fairly high sensitivity (74.7%) and specificity (88.9%)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Accuracy
    - Sensitivity highest for respiratory syndrome (74.7%)
      - GI sensitivity=72.0%
    - Specificity: ranged from 88.9%-99.9%
    - Positive predictive value ranged from 0.1% (Neurologic) to 63.8% (respiratory)
  - Timeliness: For the most common syndromes (respiratory and GI), the mean hours between hotline call and office visit was approximately 12 hours (range for other syndromes: 8.3-50.0 hours)
    - Nurse call data provides syndromic surveillance information 4-50 hours faster than outpatient office visit data

---

29. ***Title:***
Syndromic Surveillance Using Ambulatory Electronic Health Records.

Citation:
Hripcsak G, Soulakis ND, Li L, Morrison FP, Lai AM, Friedman C, Calman NS, Mostashari F. Syndromic Surveillance Using Ambulatory Electronic Health Records. *Journal of the American Medical Informatics Association* 2009;16(3):354-361.

Summary:
- *Data source:* **Ambulatory care** electronic health records from the Institute for Family Health: 13 community health centers in New York City
- *Objectives of project:* To determine the efficacy and feasibility of using electronic health record data for syndromic surveillance purposes
- *Data extracted from records:*
  - Vital signs
  - Free text notes (all)
  - Diagnosis codes
  - Reason for visit
  - Visit encounter type (all encounter types were included even virtual and no-shows)
- *Case definitions (if applicable):*
  - ILI
    - 2 days-2 weeks onset period
    - Fever over 99 degrees F in addition to at least one of the following:
      - Constitutional (fatigue, myalgia, headache, malaise)
      - Respiratory tract (cough—nonproductive, sore throat)
      - Exclude: mainly GI symptoms, sneezing/runny nose
  - GI
    - General diarrhea or vomiting within 2 weeks
    - Exclude: obvious alternate explanations for symptoms (e.g. medication side effects)
- *Data use:* Information for health services/outbreak detection
- *Data analysis:*
  - Queries were performed on structured data ('tailored approach'=appropriate for local data analysis) and natural language

processing was performed on free text data (general approach=appropriate for global data analysis)
    - Tailored and general approach systems were compared to records from 47 NYC Emergency Departments
    - Tailored approach: looked at reason for visit, ICD-9 codes, measured temperature, respiratory rate
- *Statistics/Results:*
    - Tailored data followed identified WHO ILI pattern
    - ILI data also highly correlated with validated ED data
    - General GIID data also followed identified WHO and ED ILI pattern
    - Weak correlation between general GIID data and ED data
        - May be due to low PPV (94% versus 99% PPV for ILI)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Sensitivity of tailored system: 86% (compared to ICD-9 codes for influenza)
    - ILI General approach (including at least one respiratory AND at least one constitutional symptom AND fever)
        - Sensitivity: 100%
        - Specificity: 98.6%
        - Positive predictive value: 23.4%
        - Negative predictive value: 100%
    - GIID General approach (excluding IBS, laxative use, metformin use, chronic diarrhea AND excluding migraine (vomiting))
        - Sens: 92.9%
        - Spec: 95.9%
        - PPV: 14.1%
        - NPV: 99.9%
    - Completion rate: 99%

---

30. ***Title:***
Integrating Clinical Practice and Public Health Surveillance Using Electronic Medical Record Systems.

Citation:
Klompas M, McVetta J, Lazarus R, Eggleston E, Haney G, Kruskal BA, Yih WK, Daly P, Oppedisano P, Beagan B, Lee M, Kirby C, Heisey-Grove D, DeMaria A, Platt R. Integrating Clinical Practice and Public Health Surveillance Using Electronic Medical Record Systems*. American Journal of Preventive Medicine.* 2012;42(6):S154-S162.

Summary:
- *Data source:* **Ambulatory care** electronic medical records; Electronic Record Support for Public Health (ESP) system
    - Atrius Health records: over 700,000 patients in Massachusetts
- *Objectives of project:* To examine the output/results of the ESP system in extracting relevant primary care records and sending them to public health for analysis
- *Data extracted from records:*
    - Patient demographics
    - Vital signs
    - Diagnosis codes
    - Test orders

    - Test results
    - Medication prescriptions
    - Allergies
    - Social history
    - Provider contact information
- *Case definitions (if applicable):*
    - Influenza-like illness
        - National Bioterrorism Syndromic Surveillance Demonstration Program definition is used (based on ICD-9 codes, vital signs)
    - Diabetes
        - Diagnosis codes, laboratory tests, prescriptions all used
- *Data use:* Information for health services/outbreak detection/situational awareness
- *Data analysis:*
    - Data are used either as individual case reports (when necessary) or as an aggregated data set presented to public health
    - Algorithms are put in place to identify notifiable diseases, syndromes and diabetes
    - Analyzed ESP results to determine the percentage of visits for ILI
- *Statistics/Results:*
    - ESP system contributes the most information to Massachusett's ILI surveillance program out of any other program in the state
    - Diabetes surveillance results are mapped by zip code and searchable by other demographic and risk factor information
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
    - Timeliness: sent automatically from primary care physician's record to public health (no time frame noted)
    - Validity: use validated CDC definition for ILI

---

31. ***Title:***
Using automatic medical records of rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection.

Citation:
Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automatic medical records of rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *BMC Public Health.* 2001;1:9.

Summary:
- *Data source:* **Ambulatory care** health plan records from 1996-1999; Harvard Vanguard Medical Associates (HVMA) in Massachusetts
    - Includes routine encounters, urgent care, and telephone encounters
    - 250,000 total health plan members
    - 152,435 lower respiratory infections were identified
    - Transmittal is automated
- *Objectives of project:*
    - To assess the ability of ambulatory care records to detect lower respiratory infection outbreaks and events of public health significance
- *Data extracted from records:*
    - ICD-9-CM diagnostic codes (only first 2 ICD-9-CM codes were considered)

    - Vital signs
    - Free text encounter notes
    - Patient identifier
    - Age
    - Gender
    - Membership information (in HMO)
- *Case definitions (if applicable):* Used syndrome definition from Department of Defense ESSENCE system (lower respiratory infection syndrome ONLY)
- *Data use:* Outbreak detection
- *Data analysis:*
    - Records were included in study if either of the 2 ICD-9-CM codes fit in with DoD ESSENCE syndromic definition
    - Frequency counts for ICD-9-CM codes indicative of lower respiratory infections
    - Comparisons between detected lower respiratory infections and CDC counts of pneumonia and influenza mortality data
    - Repeat encounters for LRI were only included in analysis if they occurred more than 6 weeks after the previous encounter for LRI
- *Statistics/Results:*
    - 152,435 lower respiratory infections were recorded during HVMA study period
        - 75,747 unique members
        - mean of 2 encounters per member (range: 1-65 encounters)
    - Cough ICD-9-CM code accounted for 52.8% of lower respiratory infection designations
    - Weekly counts of LRI increased greatly during the winter season
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Validity: weekly counts of pneumonia and influenza deaths from CDC followed closely with data on lower respiratory infections from HVMA
    - Timeliness: information is available within 24 hours of patient encounter
    - Completeness: authors note that completeness is equivalent to completeness of any handwritten physician record

---

32. ***Title:***
Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events.

Citation:
Lazarus R, Kleinman K, Dashevsky I, Adams C, Kludt P, DeMaria A. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerging Infectious Diseases*. 2002;8:752-760.a

Summary:
- *Data source:* **Ambulatory care** records from eastern Massachusetts
    - Includes telephone encounters, routine encounters, urgent-care encounters (not emergency department)
    - 175,000 people are in the HMO study population
- *Objectives of project:* To assess the ability of ambulatory care electronic health records to identify clusters of illness in identified syndrome categories
- *Data extracted from records:*
    - ICD-9 diagnosis code

    - Demographic information
    - Patient address
- *Case definitions (if applicable):* 9 separate ICD-9 code syndrome categories
    - ILI secondary definition:
        - Fever of over 37.8 degrees Celsius AND
        - Cough and/or sore throat
        - Exclude new encounters within 6 weeks of prior visit (for deduplication purposes)
            - 38% of encounters were removed from records because of this criterion
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:*
    - Rates of each type of syndrome were calculated
    - Searched for aberration events/large clusters of disease
    - Data were used to look for correlations between ambulatory care and inpatient hospitalizations
    - Created generalized linear mixed model (GLMM) to identify historical rates of syndrome categories
    - Used model to detect when more events were occurring than expected
- *Statistics/Results:*
    - With 1 week lag time (ambulatory care visits occurring 1 visit before hospitalizations), results showed a correlation of 0.90
        - Highest correlation apparent with 2 weeks lag time (0.92)
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
    - Timeliness: data transfer is automatic, part of physician workflow
    - Validity: since early indicators of an outbreak are not that specific, it is useful and valid to utilize syndromic definitions to search for aberration events

---

33. ***Title:***
Using Nurse Hot Line Calls for Disease Surveillance.

Citation:
Rodman JS, Frost F, Jakubowski W. Using Nurse Hot Line Calls for Disease Surveillance. *Emerging Infectious Diseases.* 1998;4(2):329-332.

Summary:
- *Data source:* Three nurses hotlines in Milwaukee, WI, 1993 **(Ambulatory)**
    - Volume of calls were tracked as well as specific symptoms
- *Objectives of project:*
    - Evaluate effectiveness of nurse hotline calls at detecting crypto outbreak prior to other methods of surveillance
    - Find out: calls increase?
- *Data extracted from records:*
    - Answers to symptom-based questions (e.g., what symptoms are you experiencing?)
    - Patient zip code
- *Case definitions (if applicable):* Diarrhea
    - 5 or more watery stools every 6 hours
- *Data use:* Outbreak detection/situational awareness
-

- *Data analysis:*
  - Data were analyzed for call volume
  - Data were analyzed for to look at the number of calls by symptom group as well as overall call volume
- *Statistics:*
  - Overall call volume: 17-fold increase during crypto outbreak
  - Diarrhea-related calls (volume) was more than 4 standard deviations above the mean number of calls
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Timeliness: information sent automatically (timeframe not specified)
  - Accuracy: callers reporting diarrhea-related illness were from the same areas of the city experiencing the outbreak (North and NW)
- *Conclusion:* call volume may be able to predict outbreaks in real time before other surveillance types

---

34. ***Title:***
Automated detection of GI syndrome using structured and non-structured data from the VA EMR.

Citation:
South BR, Gundlapalli AV, Phansalkar S, et al. Automated detection of GI syndrome using structured and non-structured data from the VA EMR. *Advances in Disease Surveillance* 2007;4:62.

Summary:
- *Data source:* Veterans Affairs (VA) electronic medical record, October 2003-March 2004; **outpatient data**
  - Data from Salt Lake City and Maryland VA records
- *Objectives of project:* To determine the efficacy of VA EMR records in detecting GI syndrome
- *Data extracted from records:*
  - ICD-9 diagnosis codes
  - Free text index patient encounter field
- *Case definitions (if applicable):*
  - Vomiting OR diarrhea OR abdominal pain
  - Less than 7 days of illness
  - Exclude non-infectious etiology
- *Data use:* Outbreak detection
- *Data analysis:*
  - Records were randomly sampled from 2 VA health systems, Maryland and Utah
  - Total of 15,377 were randomly selected for review
  - ICD-9 codes were analyzed using ESSENCE and BioSense GI case definition
  - Free text fields were analyzed using an algorithm that mined free text searching for terms that fit with established case definition
  - Rates of GIID detection found through ICD-9 and free text fields were compared to manually detected GIID (provided through chart review)
- *Statistics/Results:*
  - 43 visits fit GIID syndrome (chart review: used as gold standard)
  - ICD-9 code analysis had higher specificity, lower sensitivity than free-text

analysis
    - Combining the two methods (free text and ICD-9 code) may provide the highest level of precision in identifying GIID
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Free text mining
        - Sens: 95%
        - Spec: 85%
    - ICD-9 code
        - Sens: 54%
        - Spec: 99%
    - Overlapping free text and ICD-9 codes
        - Sens: 51%
        - Spec: 98%
    - Either ICD-9 codes OR free text mining
        - Sens: 100%
        - Spec: 84%

---

35. ***Title:***
A national syndromic surveillance system for England and Wales using calls to a telephone helpline.

Citation:
Smith GE, Cooper DL, Loveridge P, Chinemana F, Gerard E, Verlander N. A national syndromic surveillance system for England and Wales using calls to a telephone helpline. *Eurosurveillance* 2006;11: 220–224.

Summary:
- *Data source:* 22 National Health Services direct call centers **(ambulatory care)**
    - Covers populations in England and Wales
    - Nurse-led call in center
    - 3% of total primary care encounters occur through this call center
- *Objectives of project:* To assess the efficacy and timeliness of gathering syndromic surveillance data through an instituted nurse call center
- *Data extracted from records:*
    - Symptoms (not specified further)
- *Case definitions (if applicable):* Monitoring occurs for 11 separate syndromes
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data are used to track weekly rates in all 11 syndrome groupings and search for anomalies in the data
    - 99.5% confidence intervals are calculated for each syndrome group (determining 99.5% upper limit for each syndrome, as a percentage of total calls to call center)
    - Poisson distribution calculated for 6 major syndromes (including cold/flu, fever, diarrhea, vomiting) to monitor for alert-level and statistically significant increases in numbers for these syndromes
- *Statistics/Results:*
    - From March 2004-February 2005: 158 alerts
        - Of these 158: 24 (14.6%) required stage 2 investigation (further action, consultation)
        - 3 (1.9%) required stage 3 investigation (reports to local public

health or NHS medical adviser)
    - Better at detecting large events (higher number of cases)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: real time data
    - Timeliness: data are sent automatically to Health Protection Agency, analyses are performed automatically and alerts and signals are completely automatic
    - Timeliness: Only system in place to monitor heat-related illness in real time during 2006 heat wave
    - Validity: One stage 3 alert occurred at the same time as the start of flu season (followed closely with other ILI surveillance detection for the same time period)

---

36. ***Title:***
Evaluation of Electronic Ambulatory Care Data for Use in the Influenza-like Illness Surveillance Network (ILINet).

Citation:
Stigi K, Baer A, Lofy K. Evaluation of Electronic Ambulatory Care Data for Use in the Influenza-like Illness Surveillance Network (ILINet). Online Journal of Public Health Informatics 2013; 5(1): 205.

Summary:
- *Data source:*
    - Ambulatory data from Washington State
    - Emergency department data from Washington State (for comparison)
- *Objectives of project:* to determine if a current definition for influenza-like illness (ILI) (used in ED settings) could be applied to ambulatory settings to identify ILI
- *Data extracted from records:*
    - ICD codes
    - Chief complaint
- *Case definitions (if applicable):*
    - For ICD codes/diagnosis: influenza; “flu”; fever + cough; fever + sore throat
    - For chief complaint: “flu”; fever + cough; fever + sore throat
- *Data use:* Syndrome validation, situational awareness, outbreak detection
- *Data analysis:*
    - Public Health Seattle King County (PHSKC) receives data automatically from 10 outpatient clinics
    - Data from August 2007-August 2012 was queried searching for ILI, as defined in Case Definitions section above
    - Looked at correlation between proportion of ILI visits in ambulatory settings and known positive influenza laboratory tests (as determined by University of Washington Virology Laboratory)
- *Statistics/Results:*
    - Three of 10 initial clinics were excluded because of insufficient data
    - Median number of weekly visits (in 7 clinics that were analyzed): 6,167
    - Correlation between proportion of ILI visits and number of positive influenza tests 0.62-0.83 (influenza tests from individual clinics)

    - Correlation between proportion of ILI visits and number of positive influenza tests from UW lab 0.85
    - Correlation between proportion of ILI visits and percentage of positive influenza tests from UW lab 0.77
    - Strongest correlation between proportion of ILI visits and number of positive influenza tests in age group 5-24 years old (correlation coefficient 0.86)
- *Data quality:*
    - 3 of 10 initially identified outpatient clinics were excluded due to insufficient data
    - No further information on data quality

---

37. ***Title:***
Evaluating Real-Time Syndromic Surveillance Signals from Ambulatory Care Data in Four States.

Citation:
Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, Kruskal BA, Kulldorff M, Leach M, Nordin J, Patton-Levine J, Puga E, Sherwood E, Shui I, Platt R. Evaluating Real-Time Syndromic Surveillance Signals from Ambulatory Care Data in Four States. *Public Health Reports* 2010;125:111-125.

Summary:
- *Data source:* **Ambulatory care** electronic health record systems in 4 states:
    - California
    - Massachusetts
    - Minnesota
    - Texas
    - Covers one million ambulatory care patients
- *Objectives of project:* To determine how well ambulatory care syndromic surveillance systems can detect outbreaks compared to traditional surveillance systems
- *Data extracted from records:*
    - ICD-9-CM diagnostic codes
- *Case definitions (if applicable):*
    - ICD-9-CM diagnostic code-based syndrome definitions developed by CDC/Dept of Defense working group (2003)
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data from the electronic health records were gathered and parsed for signals that may be of public health interest
    - Data were analyzed in near real-time (i.e., not retrospectively)
        - Aggregated data were sent daily to public health
    - Surveillance period for 3 of the sites was one year (specific month of surveillance start varied somewhat between sites)
        - Surveillance period for Massachusetts site was 4 months due to error that occurred 4 months into project
    - Data were collected and analyzed for frequency of syndromes
    - Frequency of syndromes was then compared to known, related outbreaks occurring during the same time period (picked up through traditional surveillance methods)

    - Repeat visits within same 6 week period for same time period were removed from analysis
    - Analyzed for spatial-temporal patterns as an aggregated data set
    - Used 365 days preceding study time as a historical data comparison
    - Used recurrence interval (RI) of 365 days or greater to signal statistical significance
- *Statistics/Results:*
    - 17 distinct signals (62 alerts were grouped into these 17 signals) were picked up the syndromic surveillance system
        - Only 2 of the signals picked up true clusters of verified illness (varicella)
        - No novel outbreaks (missed by traditional surveillance) were detected
    - During the same time period currently implemented traditional surveillance system detected 124 clusters of illness
    - True/plausible signals overall had (compared to false signals):
        - Higher median number of cases
        - Smaller geographic area
        - Higher recurrence interval
        - Shorter signal length (slight difference)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: Files were sent automatically each day for the previous 24 hour period
    - Accuracy: 15 false alerts were observed during the study period (36 could be expected)
    - Validity: Application across states and different health systems indicates substantial generalizability of study results

---

38. ***Title:***
Telephone Triage Service Data for Detection of Influenza-Like Illness.

Citation:
Yih WK, Teates KS, Abrams A, Kleinman K, Kulldorff M, Pinner R, Harmon R, Wang S, Platt R. Telephone Triage Service Data for Detection of Influenza-Like Illness. *PLoS ONE*. 2009;4(4):e5260.

Summary:
- *Data source:* Nurse telephone triage call data **(ambulatory)** gathered from the 17 US states that have more than 500,000 inhabitants; October 2004-April 2005 influenza season
- *Objectives of project:* To monitor the efficacy of syndromic surveillance through a telephone call center and to compare the syndromic system to CDC weekly positive influenza specimens (viral isolates) and sentinel providers' ILI rates (as percentage of total visit)
- *Data extracted from records:*
    - Demographic information
        - Patient zip
    - List of guidelines consulted (guidelines provided information to nurses for various illnesses—guidelines related to syndromes of interest were flagged and those records were extracted for study)
- *Case definitions (if applicable):* Influenza-like illness (applicable to sentinel

providers' ILI rates ONLY not to nurse triage call data):
    - Temperature of 100 degrees farenheit or higher
    - Cough and/or sore throat
    - Exclude: anyone with other known cause of symptoms
- *Data use:* Situational awareness/outbreak detection
- *Data analysis:*
    - Data were used to track ILI and to compare syndromic surveillance-detected ILI to other known morbidity measures for ILI
    - State-level analyses were conducted
    - CDC data and sentinel provider data were compared to respiratory syndrome data (all 3 sources were grouped weekly)
    - Pearson correlations were calculated
- *Statistics/Results:*
    - Median correlation between CDC viral isolates and sentinel data was highest (0.80)
        - 12/17 states (70.6%) saw correlations >/=0.75
    - Next highest correlation: telephone triage data and sentinel provider percentage rates (median correlation=0.74; range 0.34-0.89)
    - Median correlation between telephone triage data and CDC viral isolates was 0.65 (range 0.35-0.85)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness: data were sent daily for previous days' telephone encounters
    - Accuracy: poor agreement between more specific measure (viral isolates) and telephone triage center; much better agreement between more sensitive measure (sentinel provider percentage rates) and telephone triage center
    - Completeness: 753/1291 sentinel providers in 17 study states regularly submitted reports (58.3%)
        - "regularly submitted reports"=at least 16 of the weeks of influenza season (at least 50% of the time)
    - Validity may be low because triage center coverage was not spread evenly across all geographic areas in the 17 states represented
        - Results are more valid in some states than in others
    - Accuracy: ILI definition was very broad and included things such as asthma, sinus pain (fever was not required for telephone data)

39.*Title:*
Automated Surveillance of Outpatients with Pneumonia: A Performance Evaluation.

Citation:
Zheng H, Siddiqui T, DeLisle S. Automated Surveillance of Outpatients with Pneumonia: A Performance Evaluation. *Online Journal of Public Health Informatics.* 2013;5(1):224.

Summary:
- *Data source:* Outpatient records from VA Maryland Healthcare System
- *Objectives of project:* to determine if tracking all acute respiratory infections (ARI) for influenza surveillance is worthwhile or whether it should be focused only on pneumonia
- *Data extracted from records:*
    - ICD-9 codeset related to ARI

    - Chest imaging
    - Text analysis
- *Case definitions (if applicable):*
    - ARI: ICD-9 code from ARI codeset
    - Pneumonia: ARI ICD-9 code + chest imaging + imaging text analysis indicating pneumonia diagnosis
- *Data use:* situational awareness (determining severity of influenza outbreaks)
- *Data analysis:*
    - Time series analysis performed by applying both ARI and Pneumonia case definitions to diagnoses, imaging, and text analysis
    - Calculated “detection delay”—average amount of time from case to first alarm
    - Calculated false alarm rate (number of unique false alarms during study year)
- *Statistics/Results:*
    - Detection delay is lower when looking at pneumonia diagnoses not general ARIs
    - General conclusion of data: influenza surveillance targeted at pneumonia is more effective than those that monitor for all ARIs
- *Data quality:*
    - Sensitivity: detection delay is lower at any given false alarm rate when looking at pneumonia versus ARI

RETURN TO TABLE OF CONTENTS

Potential Uses Utilizing Future Data Elements

This section, *Potential Uses Utilizing Future Data Elements,* describes research and practice in using specific, infrequently used, data from clinical settings for syndromic surveillance. For example, articles summarizing the use of laboratory data elements are in this section. These data elements are used in limited jurisdictions and research settings, but their integration is not yet widespread.

40. ***Title:***
Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks.

Citation:
Carnevale RJ, Talbot TR, Schaffner W, Block KC, Daniels TL, Miller RA. Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks. *Journal of the American Medical Informatics Association* 2011;18:466-472.

Summary:

- *Data source:* Microbiologic data from **inpatient, ambulatory,** and **emergency department**
  - Sample information was from 2001-2006
  - Vanderbilt-affiliated hospitals (Tennessee)
- *Objectives of project:* To determine the utility of syndromic surveillance data to detect hospital-acquired and based infection outbreaks
- *Data extracted from records:*
  - Daily counts of each positive microbiologic test
- *Case definitions (if applicable):* N/A
- *Data use:* Outbreak detection
- *Data analysis:*
  - Phase 1: configured 4 algorithms
  - Phase 2: algorithms were reviewed and noted outbreaks were deemed “probable” or “possible”
  - Phase 3: evaluation
  - Possible outbreaks were compared to known outbreaks
  - Algorithms were assessed for their ability to identify the found outbreaks
- *Statistics/Results:*
  - Phase 1: Decay rate of 0.3, alerting threshold of 5 cases were set
  - Phase 2: 257 alerts were found
    - 17 alerts were agreed upon by expert review to be candidate outbreak clusters
    - 12 additional outbreaks were added (3rd reviewer found outbreak or infection control archival data identified an outbreak)
    - Total: 29 candidate outbreaks
    - Clusters from inpatient significantly more likely to be recorded as candidate outbreaks (chi square p-value=0.002)
  - Phase 3: 2 actual outbreaks of C. difficile occurred during time of study preparation
    - Algorithms were implemented as a retrospective evaluation tool
    - Both actual outbreaks were detected, 39 other signals also occurred (were not evaluated)
- *Data quality (e.g., timeliness, accuracy, validity, completion)*

- Accuracy (of each of 4 algorithms)
  - Positive predictive value: range 5.3%-29%
  - Sensitivity range: 21%-31%

---

41. ***Title:***
Increased antiviral medication sales before the 2005-06 influenza season—New York City.

Citation:
Centers for Disease Control (CDC). Increased antiviral medication sales before the 2005-06 influenza season—New York City. *Morbidity and Mortality Weekly Report* 2006;55(10):277-279.

Summary:
- *Data source:* New York State Department of Health monitoring of antiviral **medication** sales paid by Medicaid AND New York City Department of Health and Mental Hygiene monitoring of anitiviral medication sales by a large retail pharmacy chain
- *Objectives of project:* To identify if pharmacy sales for antiviral medications can be used as an indicator of influenza outbreaks prior to their identification through other methods
- *Data extracted from records:*
  - Medication category
  - Patient postal code
  - Age group
  - Sex
- *Case definitions (if applicable):*
  - Only included oseltamivir, zanamivir, rimantadine in antivirals category: amantadine was excluded because it can also be used to treat Parkinson Disease
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:* Pharmacy records were compared to WHO and NREVSS lab reports to CDC
- *Statistics:*
  - Average of 29,664 medications reported/paid for by Medicaid each day
  - Spike appeared in data 7 weeks prior to influenza outbreak in Fall 2005; 1 week prior to influenza outbreak in Fall 2004
  - Study indicates that pharmacy sales may be a more timely indicator of influenza outbreaks than national reporting, lab-based systems
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
  - Completion/timeliness: 95% of medications paid for by Medicaid are reported within 1 day of the medication sale
  - Medicaid provides medication coverage to 34% of NYC residents
  - Timeliness: NYCDOHMH receives pharmacy files every day (contains information on all medications sold the previous day)

---

42. ***Title:***

Patterns in influenza medication use before and during the 2009 H1N1 pandemic, Vaccine Safety Datalink Project, 2000-2010.

Citation:

Greene SK, Shay DK, Yin R, McCarthy NL, Baxter R, Jackson ML, Jacobsen SJ, Nordin JD, Irving SA, Naleway AL, Glanz JM, Lieu TA. Patterns in influenza medication use before and during the 2009 H1N1 pandemic, Vaccine Safety Datalink Project, 2000-2010. *Influenza and Other Respiratory Viruses*. 2012; DOI: 10.1111/j. 1750-2659.2012.00390.x.

Summary:

- *Data source:* 8 U.S. **ambulatory, emergency department and inpatient** records from medical care organizations that were participating in the Vaccine Safety Datalink Project
    - HealthPartners Research Foundation (Minnesota) January 1, 2000-June 30, 2010
    - Marshfield Clinic Research Foundation (Wisconsin) January 1, 2000-June 30, 2010
    - Kaiser Permanente of Northern California January 1, 2000-June 30, 2010
    - Group Health Cooperative (Washington) January 1, 2001-June 30, 2010
    - Kaiser Permanente of Southern California, January 1, 2006-June 30, 2010
    - Harvard Vanguard Medical Associates and Harvard Pilgrim Health Care (Massachusetts) January 1, 2007-June 30, 2010
    - Kaiser Permanente of Colorado, January 1, 2007-June 30, 2010
    - Kaiser Permanente Northwest (Oregon), January 1, 2007-June 30, 2010
- *Objectives of project:* To monitor the rate of antiviral medication dispersal before and during the H1N1 pandemic in 2009
- *Data extracted from records:* antiviral medication dispensings; specific details included:
    - Number of days supply of medication
    - Date of dispensing
    - Number of units of medication
    - Medications monitored: oseltamivir, zanamivir, rimantadine, amantadine
- *Case definitions (if applicable):* Patients were only included in the analysis if they were enrolled in their medical care organization (MCO) for 7 days before the medication was dispensed to 42 days after the medication was dispensed. Otherwise the medication was determined not to be for “treatment” purposes (could be for stockpiling or prophylaxis)
- *Data use:* Outbreak detection/situational awareness
- *Data analysis:*
    - The data was used to track the use of antiviral medications in an ambulatory setting
    - Used as a way of tracking ILI→ medication orders as a proxy for the prevalence of influenza in the population
    - Data for antiviral dispensings for treatment were compared to CDC guidelines (to monitor for adherence to drug resistance avoidance guidelines) and to weekly percentages of influenza tests that were positive in the region.

  Statistics:
    - 66,698 antiviral medication dispersals were made
        - 59.4% were determined to be for treatment purposes

    - 1/3 of patients who did not receive an influenza diagnosis were treated with antiviral medication, indicating that monitoring vaccination may be a useful proxy measure for influenza trends
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - No information in article re: data quality

---

43. ***Title:***
Chronic Disease and Disasters: Medication Demands of Hurricane Katrina Evacuees.

Citation:
Jhung MA, Shehab N, RohrAllegrini C, Pollock DA, Sanchez R, Guerra F, Jernigan DB. Chronic Disease and Disasters: Medication Demands of Hurricane Katrina Evacuees. *American Journal of Preventive Medicine.* 2007;33(3):207-210.

Summary:
- *Data source:* Healthcare encounter records from Hurricane Katrina (August, 2005) evacuees
    - 4 main evacuee centers: **ambulatory and prescription data**
    - Records from 4-31 days after hurricane made landfall
    - Evacuees were relocation to San Antonio, TX from Louisiana and Mississippi
- *Objectives of project:* To determine the relationship between medication demands and pharmaceutical supplies following a disaster
- *Data extracted from records:*
    - Chief complaint
    - Pharmacy records from local retail pharmacies, specific evacuee centers (only records from evacuees were included)
        - Prescription
        - Over-the-counter medications
- *Case definitions (if applicable):*
    - Medications were classified into “acute” or “chronic” category
- *Data use:* Information for public health services
- *Data analysis:*
    - Rates of medication for acute and chronic illness categories were calculated separately
    - Total prescription numbers were tallied
- *Statistics/Results:*
    - Median number of daily encounters: 218 (5% of evacuee population)
    - 14,719 Defined Daily Doses of medication were dispensed
    - 67% of categorized healthcare encounters were for acute conditions
    - 15% of categorized healthcare encounters were for chronic conditions
        - 68% of Defined Daily Doses were for chronic disease
    - Authors note that there is evidence that chronic disease burden post-disaster is substantial
- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Completion: 11% of retail pharmacy records were utilized in analysis
    - Timeliness: not specified

---

44. ***Title:***
Code-based Syndromic Surveillance for Influenza like Illness by International Classification of Diseases, Ninth Revision.

Citation:
Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang J, Pavlin JA. Code-based Syndromic Surveillance for Influenzalike Illness by International Classification of Diseases, Ninth Revision. *Emerging Infectious Diseases*. 2007;13(2):207-216.

Summary:
- *Data source:* ESSENCE **outpatient** data (includes primary care and emergency departments)
    - Data gathered from military treatment facilities (MTFs)
        - Only US Air Force MTFs data were included because records could be linked to lab results
    - Over 300,000 weekly visits on average
    - Study period: June 2002-June 2004
- *Objectives of project:* To determine the correlation between ICD-9 based influenza-like illness classification and actual rates of influenza, as determined by laboratory results obtained from the same source population
- *Data extracted from records:*
    - ICD-9 codes
- *Case definitions (if applicable):*
    - 29 ICD-9 codes were considered to indicate influenzalike illness, including (for initial anaysis):
        - Fever
        - Acute respiratory infection
        - Unspecified viral illness
    - After matched analysis, 14 total ICD-9 codes were determined to correctly indicate influenza
- *Data use:* Information for health services/outbreak detection/situational awareness
- *Data analysis:*
    - Records were gathered and analyzed for rates of influenza like illness
    - Lab results data from MTFs (especially sentinel sites, which are required to send weekly samples during influenza season)
        - Data gathered from patients fitting definition of ILI:
            - Fever of 100.5 degrees Fahrenheit or higher AND
            - Cough or sore throat
    - Lab results data were matched with outpatient care record to search for correlation between a positive influenza test result and influenza like illness identified through syndromic means
        - Lab results linked to a patient encounter (and ICD-9 code) then had a single ICD-9 code selected (first: diagnosis most related to influenza; if multiple, then most severe was selected first)
    - Data were also compared to Department of Defense Global Influenza Surveillance Program rates (this analysis did not include linked data, just looked at rate trends)
- *Statistics/Results:*
    - Matched analysis: 6,236 specimens were matched with ICD-9 codes
    - ICD-9 codes that showed high correlation with positive lab results usually also had high signal:noise ratio
    - 14 ICD-9 codes were determined to best indicate influenza

- Weekly temporal trends of ILI had statistically significant correlations with all 23 ICD-9 codes (0.72, $p<0.0001$), narrowed list of 14 ICD-9 codes (0.71, $p<0.0001$) and an even smaller list of 4 ICD-9 codes (0.86, $p<0.0001$)
    - Small group (4):
        - Infectious upper respiratory, multiple sites, acute NEC
        - Influenza with pneumonia
        - Influenza with respiratory manifestation NEC
        - Influenza with manifestation NEC
- CDC data were most closely correlated with the smaller group of 4 ICD-9 codes (correlations ranged from 0.87 in 2002-2003 season to 0.99 in 2003-2004 season, all 3 influenza seasons showed statistically significant correlations)
- Conclusion: both large group and small group ICD-9 codes are effective ways of monitoring for influenza like illness in a more timely fashion than via lab results

- *Data quality (e.g., timeliness, accuracy, validity, completion)*
    - Timeliness:
        - Data sent automatically, reporting time ranges from 1-4 days on average
        - 80% of all data received within 4 days
        - Data pulled from central database to ESSENCE server every 8 hours
        - Syndromic data is available much more quickly than lab data (because of unavoidable lag due to lab analysis time)

---

45. ***Title:***
A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions.

Citation:
Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health 2008;98(2):344–350.

Summary:

- *Data source:* Data retrieved from **inpatient** and **ambulatory** sources in Marion County, Indiana from January-March 2001
    - Public health department records
    - Health information exchange
    - Hospital infection control departments
- *Objectives of project:* To determine the completeness and timeliness of electronic laboratory reporting (ELR) in comparison to paper-based reporting of notifiable diseases
- *Data extracted from records:*
    - Laboratory results
    - Medical record number
    - Name
    - Date of birth
    - Gender
    - Street address (patient)

        - City (patient)
        - Zip code (patient)
        - Telephone number (patient)
        - Specimen collection date
        - Specimen type
        - Physician name
        - Physician street address
        - Physician city
        - Physician state
        - Physician zip code
        - Physician telephone number
    - *Case definitions (if applicable):*
        - ELR: Positive laboratory results for 55 laboratory notifiable conditions
        - Paper records: Case reported to public health as notifiable condition
    - *Data use:* Inform public health services about the benefits of electronic laboratory reporting
    - *Data analysis:*
        - Cases of notifiable disease were linked using medical record number, name, date of birth
        - Compared unique cases to the direct reporting (via paper records) of these diseases to Marion County Health Department
        - Measured completeness in fields required by public health for analysis
    - *Statistics/Results:*
        - 4,785 unique cases of notifiable diseases were reported
        - ELR reported 4.4 times the number of cases of paper reporting (4,625 vs. 944)
        - 818 cases were reported in both paper records and ELR
    - *Data quality (e.g., timeliness, accuracy, validity, completion)*
        - Completeness: notifiable-condition database received lab results from 9/13 county hospitals (60% of lab results in the study were from these 9 hospitals)
        - Timeliness: ELR identified cases on average 7.9 days earlier than paper reporting
        - Completeness:
            - Data from 10 (of 18) of the fields was filled in more often in the ELR reports than in the paper reports (5 of the fields were filled more often in paper reporting than ELR reporting)

---

46. ***Title:***
Ambulatory-care diagnoses as potential indicators of outbreaks of gastrointestinal illness—Minnesota.

Citation:
Yih WK, Abrams A, Danila R, Green K, Kleinman K, Kulldorff M, et al. Ambulatory-care diagnoses as potential indicators of outbreaks of gastrointestinal illness—Minnesota. *MMWR Morb Mortal Wkly Rep*. 2005;54(Suppl):157-62.

Summary:
- *Data source:* **Ambulatory care** records from a Minnesota health plan, 2001-2003 (2 years)
    - 240,000 members

- *Objectives of project:* To identify the efficacy of ambulatory care records to detect a gastrointestinal illness outbreak
- *Data extracted from records:* (not explicitly stated, but these elements are referenced in the article)
    - Visit date
    - Zip code
    - ICD-9 diagnostic codes
- *Case definitions (if applicable):*
    - Used syndrome definitions from CDC/Dept of Defense working group
- *Data use:* Outbreak detection
- *Data analysis:*
    - Data from the ambulatory care records were compared to known morbidity frequencies from the Minnesota Department of Health (gathered from other surveillance methods)
    - Recurrence intervals were calculated (RI)
        - RI >/= 2 years was considered highly significant
        - RI>/= 2 weeks were also considered
    - Frequency counts were compared to verified outbreaks
- *Statistics/Results:*
    - 110 outbreaks were identified by MN Dept of Health as verified outbreaks
    - 1.3 million person-years of data were received
    - 1 signals identified with RI>/=2 years
    - 58 signals identified with RI>/= 2 weeks
    - Number of signals was not statistically significant (p=0.22 for RI>/=2 weeks: one-sided test)
    - As time frame increased positive predictive value increased, sensitivity decreased (looking at efficacy of syndromic system to detect known outbreaks)
- *Data quality (e.g., timeliness, accuracy, validity, completion):*
    - Timeliness: near real-time (illness counts are sent every day)
    - Completeness: covered 8% of Minneapolis-St. Paul area
    - Accuracy (of syndromic system to known outbreaks concordance):
        - RI >/= 2 years
            - Sensitivity: 1%
            - Positive predictive value (PPV): 100%
        - RI>/=2 weeks
            - Sensitivity: 29%
            - Positive predictive value: 13%

RETURN TO TABLE OF CONTENTS