# Evaluating Real-Time Syndromic Surveillance Signals from Ambulatory Care Data in Four States

W. Katherine Yih, PhD, MPH[a]
Swati Deshpande, PhD[b]
Candace Fuller, MPH[c]
Dawn Heisey-Grove, MPH[d]
John Hsu, MD, MBA, MSCE[e]
Benjamin A. Kruskal, MD, PhD, MS[f]
Martin Kulldorff, PhD[a]
Michael Leach, MPH[b]
James Nordin, MD, MPH[g]
Jessie Patton-Levine[h]
Ella Puga, MPH[h]
Edward Sherwood, MD[i]
Irene Shui, MPH[a]
Richard Platt, MD, MS[a]

## SYNOPSIS

**Objectives.** We evaluated a real-time ambulatory care-based syndromic surveillance system in four metropolitan areas of the United States.

**Methods.** Health-care organizations and health departments in California, Massachusetts, Minnesota, and Texas participated during 2007–2008. Syndromes were defined using International Classification of Diseases, Ninth Revision diagnostic codes in electronic medical records. Health-care organizations transmitted daily counts of new episodes of illness by syndrome, date, and patient zip code. A space-time permutation scan statistic was used to detect unusual clustering. Health departments followed up on e-mailed alerts. Distinct sets of related alerts ("signals") were compared with known outbreaks or clusters found using traditional surveillance.

**Results.** The 62 alerts generated corresponded to 17 distinct signals of a potential outbreak. The signals had a median of eight cases (range: 3–106), seven zip code areas (range: 1–88), and seven days (range: 3–14). Two signals resulted from true clusters of varicella; six were plausible but unconfirmed indications of disease clusters, six were considered spurious, and three were not investigated. The median investigation time per signal by health departments was 50 minutes (range: 0–8 hours). Traditional surveillance picked up 124 clusters of illness in the same period, with a median of six ill per cluster (range: 2–75). None was related to syndromic signals.

**Conclusions.** The system was able to detect two true clusters of illness, but none was of public health interest. Possibly due to limited population coverage, the system did not detect any of 124 known clusters, many of which were small. The number of false alarms was reasonable.

[a]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

[b]San Mateo County Health Department, San Mateo, CA

[c]Infectious Disease Epidemiology Prevention and Control Division, Minnesota Department of Health, St. Paul, MN

[d]Bureau of Communicable Disease Control, Massachusetts Department of Public Health, Boston, MA

[e]Kaiser Foundation Research Institute, Oakland, CA

[f]Harvard Vanguard Medical Associates, Boston, MA

[g]HealthPartners Research Foundation, Minneapolis, MN

[h]Austin/Travis County Health and Human Services Department, Austin, TX

[i]Central Texas Veterans Health Care System, Temple, TX

Address correspondence to: W. Katherine Yih, PhD, MPH, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Ave., 6th Floor, Boston, MA 02215; tel. 617-509-9822; fax 617-859-8112; e-mail <katherine_yih@harvardpilgrim.org>.

The past decade has seen the development of dozens of syndromic surveillance systems, both domestically and internationally. Although a crucial impetus was U.S. governmental concern about bioterrorism, an important goal of syndromic surveillance has been the timely detection of naturally occurring disease outbreaks to augment health departments' routine surveillance methods.

The need to evaluate these systems was recognized early,[1,2] particularly in light of the diversion of resources to develop them and skepticism from some critics in the public health community.[3] In the United States, most syndromic surveillance systems use emergency department (ED) encounter data,[4] and a number of evaluations of these ED-based systems have been published at the city,[5,6] county,[7] and state[8] levels. While some disease outbreaks were detected, these studies found no outbreaks of public health significance that were missed by existing traditional surveillance systems.

Use of electronic medical record data from ambulatory care settings for syndromic surveillance, which might capture more episodes of illness—and possibly in earlier stages—is less common. Analyses using diagnostic codes from ambulatory care episodes and from ED episodes from the same overall patient population yield different sets of signals,[9,10] and one data source may be better than the other for detecting some outbreaks of naturally occurring disease.

The multisite, ambulatory care-based system described in this article was developed over the course of several years, with an original goal of detecting and reporting unusual geographic and temporal clusters of ambulatory care encounters for acute illness that might represent the initial manifestations of a bioterrorism event. Its establishment required major effort on the part of informaticians, programmers, statisticians, epidemiologists, public health officials, and administrative support staff. The development work included (1) designing a system to allow transmission of aggregate data to a central data repository, while leaving patient-level data at the respective health-care organization in a form easily accessible in the event of a signal; (2) supporting the use of data-processing software by multiple remote users; (3) evaluating different statistical techniques, models, and parameter settings to adjust for the natural variation in disease occurrence; (4) designing a website to display results and data to participants; and (5) automating alerting.[11–13] This collaboration produced a number of retrospective analyses, including comparisons of signals with known gastrointestinal (GI) illness outbreaks[14] and with indicators from an existing influenza surveillance system.[15]

System evaluation during real-time operation is a more definitive approach than retrospective evaluation, as it involves investigation of freshly identified clusters, likely to be carried out more assiduously and with greater access to supplemental surveillance information than investigation of old clusters. We describe results of such an evaluation of our real-time ambulatory care-based syndromic surveillance system in four metropolitan areas during up to one year of surveillance. The system was evaluated in terms of ability to detect disease outbreaks or clusters worthy of health department attention, number of false alarms, and time expended investigating signals.

## METHODS

### Population and participating organizations

The population covered by the surveillance system was approximately one million. The four locations, with their data providers and collaborating health departments, were greater Boston, Massachusetts (Harvard Vanguard Medical Associates and Massachusetts Department of Public Health); greater Minneapolis–St. Paul, Minnesota (HealthPartners and Minnesota Department of Health); greater Austin, Texas (Austin Diagnostic Clinic and Austin Regional Clinic, Austin/Travis County Health and Human Services Department); and San Mateo County, California (Kaiser Permanente Northern California and San Mateo County Health Department). Table 1 shows, for each surveillance area, the population in the surveillance system and the proportion it comprised of the total population.

### Surveillance period

For three of the sites, the surveillance period was one year, including most of 2007: California's period was from January 22, 2007, to January 21, 2008; the surveillance period in Minnesota and Texas was from February 8, 2007, to February 7, 2008. In Massachusetts, an error was inadvertently introduced after four months of real-time surveillance operation, which affected the types of encounters captured in the source data; this led us to shorten the surveillance period to the first four months: March 27 to July 26, 2007.

### Syndromes

Nine syndromes defined by International Classification of Diseases, Ninth Revision Clinical Modification (ICD-9-CM) codes were tracked: respiratory, influenza-like illness (ILI), upper GI, lower GI, hemorrhagic, lesions, lymphadenopathy, neurologic, and rash. All except for ILI comprised a subset of the full code lists developed for these syndromes by a Centers for

Disease Control and Prevention (CDC)/Department of Defense working group in 2003.[16] The full code lists were pared down to more specific definitions by the Massachusetts Department of Public Health, and the ILI definition was created in collaboration with that health department.

CDC defines ILI as a fever of at least 100°F with cough or sore throat, in the absence of other confirmed diagnoses; this definition is used to determine the percentage of visits to sentinel providers that are for ILI, one of CDC's influenza surveillance indicators.[17] Our ILI definition required (1) a fever of at least 100°F or, if there was no measured temperature, the ICD-9-CM code for fever, and (2) at least one of 31 respiratory codes. These respiratory codes overlapped with but were not a subset of the respiratory syndrome codes. For example, codes for tracheitis and pneumonia were in common, but the ILI definition included various kinds of acute viral upper respiratory infection, influenza, and cough, which the respiratory illness definition did not. Counts for ILI were divided into three age groups: 0–4, 5–12, and >12 years of age; these were handled as separate "syndromes" for purposes of signal detection, although they differed only in terms of age group. The syndrome definitions are available from the authors upon request.

### Data processes

The raw data came from electronic medical records, in which primary care providers had recorded ICD-9-CM codes during the routine delivery of care. Data managers at each site set up daily extraction of data conforming to uniform file specifications and installed data-processing software downloaded from the project's password-protected website. After initial setup, data processing was fully automated. Each night, the installed software identified from electronic medical records the day's encounters with ICD-9-CM codes of interest, assigned encounters to syndromes, ignored repeat visits occurring within six weeks of a visit for a previously identified episode of illness in the same syndrome, and assigned the remaining new episodes of illness to the zip codes where the affected people lived. Each site automatically transmitted the resulting aggregate data—counts of episodes by syndrome, date, and zip code—to a central data repository on a daily basis. Patient-level data remained at their respective sites, including line lists from which the relevant cases for chart review could be identified in the event of an alert. At the central data repository, the data were analyzed for unusual clustering of cases in time and space (method discussed later in this article), again, in a largely automated way. These procedures have been described more fully elsewhere.[11,12,18,19]

### Terminology

We use the following terminology:

- An *alert* is the notification that went to the health department when a statistically significant group of cases (clustered in space and time) was detected. Sometimes there were multiple alerts on consecutive days for the same syndrome at

**Table 1. Health-care organizations, health departments, surveillance areas, and populations included in an evaluation of a syndromic surveillance system, 2007–2008**

| State | Data-providing health-care organization(s) | Collaborating health department | Surveillance area | Population served by health-care organization within surveillance area | Proportion of surveillance area's total population included (percent) |
|---|---|---|---|---|---|
| Massachusetts | Harvard Vanguard Medical Associates | Massachusetts Department of Public Health | Greater Boston area[a] | 275,000 | 8 |
| Minnesota | HealthPartners | Minnesota Department of Health | Greater Twin Cities area | 285,000 | 10 |
| Texas | Austin Diagnostic Clinic | Austin/Travis County Health and Human Services Department | Austin and Travis County | 91,000 | 10 |
| Texas | Austin Regional Clinic | Austin/Travis County Health and Human Services Department | Austin and Travis County | 180,000 | 20 |
| California | Kaiser Permanente Northern California | San Mateo County Health Department | San Mateo County | 178,000 | 25 |

[a]The surveillance area was defined as all towns for which at least 5% of the residents used Harvard Vanguard Medical Associates, plus the adjacent towns. This set forms a contiguous area consisting of metropolitan Boston, suburbs and, in the outer ring, semirural areas.

approximately the same geographical location, with many cases in common.

- A *signal* is a group of such overlapping alerts, representing a single potential disease outbreak.
- *Outbreak* and *cluster* are used somewhat interchangeably, and often together, to refer to a plausibly related set of cases recognized by a health department, although *cluster* occasionally means a merely statistically identified group of cases found by our surveillance system, in which case we make that explicit.

### Statistical signal detection method

To detect unusual groupings of cases, we applied the space-time permutation scan statistic[20] to each organization's data, so data from the two Austin health-care organizations were analyzed separately. This nonparametric method required minimal assumptions and adjusted for any purely spatial and purely temporal variation. In other words, if a certain zip code was associated with consistently more visits than others or there were uniform increases or decreases in illness across a data-providing organization's whole catchment area, no signals would result.

As parameter settings, we used the previous 365 days as historical data, a maximum scanning window of 14 days (allowing for detection of outbreaks lasting between one and 14 days), a maximum geographic circle size of 50% of the population, and 9,999 Monte Carlo replications. If, during a single daily statistical analysis, multiple geographically overlapping groups of cases were detected for a particular syndrome, only the most unusual ("primary") one was reported.

The measure of statistical aberration used was the recurrence interval (RI), the inverse of the *p*-value.[21] The criterion for an alert was RI≥365 days, meaning that the event in question would be expected to occur by chance alone no more often than once a year for that particular syndrome and health organization. This alerting criterion, together with the fact that 9,999 Monte Carlo replications were implemented in the analyses, meant that the range of possible RIs for any statistically identified cluster was 365–10,000 days. We also required that there be at least three cases for an alert to be sent. We conducted the statistical analyses using SaTScan™ software.[22]

### Alerting

The detection of a statistically significant cluster during the daily scanning generated an automated alert by e-mail from the central data repository to designated staff at the respective health department. Information in the alerts included date(s), town/zip code(s) of patient residence, syndrome, number of cases by date and zip code, relative risk, RI, and a unique alert identification number. Alerts overlapping in time and space with previous alerts were clearly labeled (e.g., "This alert has been identified to overlap with alert(s) 78289 . . .").

Minnesota alerts for upper and lower GI signals were sent to a foodborne illness specialist at the Minnesota Department of Health. Minnesota alerts for other syndromes went to the clinical responder at HealthPartners Research Foundation. Each of the other health departments received all alerts for its respective jurisdiction.

### Response

The response to alerts depended on the judgment of health department staff. Sometimes no follow-up was considered necessary. Often, the health department contacted a designated clinical responder at the health-care organization whose data had generated the alert, and this responder accessed the automatically created line lists for each day of the alert to provide specific diagnoses and demographic information on the cases involved in the alert. Medical records, whose numbers were provided in the line lists, were occasionally reviewed by the clinical responder at the health-care facility or by health department personnel.

### Evaluation

Details contained in the alerts were automatically recorded by the system's central data repository; health department staff reported the time expended investigating each alert and their findings in Microsoft® Excel. (Investigation time on the part of clinical responders at the health-care organizations was not recorded.) Data elements collected for known outbreaks or clusters included syndrome, pathogen if known, date(s), location(s), case counts, and free-text comments. We included in the count of known outbreaks/clusters all those in which missing data elements made eligibility unknown (eligibility criteria are listed later in this article) or prevented us from identifying duplicate records. This inclusive approach may have led to a slightly inflated number of clusters, but any inaccuracy in this regard would not have affected our conclusions.

### Traditional surveillance

The outbreaks or clusters found by traditional surveillance included in this analysis were those that (*1*) presented as at least one of the syndromes under surveillance, (*2*) warranted an investigation by the health department, (*3*) occurred at least partially within the

surveillance area and surveillance period, (4) consisted of at least two plausibly linked cases, and (5) did not occur in institutions likely to have their own health-care services, such as long-term-care facilities, prisons, or colleges/residential schools. These criteria were established a priori.

## RESULTS

Overall, during 12 months of surveillance in three areas and four months of surveillance in the fourth, there were 62 alerts grouped into 17 distinct signals, with a median of two alerts per signal (range: 1–16). There was at least one signal in each syndrome type except lymphadenopathy. Taking as representative of a multi-alert signal the constituent cluster with the highest RI and, if there were ties, the highest number of cases and, if there were still ties, the earliest end date, the median number of cases per signal was eight (range: 3–106), the median number of zip code areas per signal was seven (range: 1–88), and the median number of days per signal was seven (range: 3–14, with 14 the maximum possible, given the 14-day scanning window). RIs ranged from 385 to 10,000 days (the maximum possible).

Details of each signal are shown in Table 2. Of the 17 signals, two were investigated and judged to be true clusters, six were investigated and the cases appeared unrelated, six were not rigorously investigated but considered plausible (i.e., thought to be consistent with seasonal outbreaks of respiratory disease and norovirus), and three were not investigated. The two rash signals, one in California and one in Massachusetts, reflected true clusters of varicella, mostly break-through illness (i.e., occurring despite vaccination and usually mild), although one of the four Massachusetts cases had been miscoded and was not a true case. As these were neither hospitalized cases nor associated with known outbreaks, epidemiologists in the respective health departments did not consider the signals worthy of investigation beyond an initial brief assessment. Regarding depth of investigation, seven signals (41%) led to chart review (one of which also involved patient interviews), five (29%) involved consultation of the line list only, and five (29%) led to neither line list nor chart review.

The eight true or plausible signals differed from the six that were determined not to be true in the following ways: more cases (median = 17.5, range: 3–106 vs. median = 8, range: 3–36), smaller geographic area (median = 2 zip code areas, range 1–88 vs. median = 9, range: 1–32), and higher RI (median = 6,000 days, range: 667–10,000 vs. median = 646, range: 385–1,667).

True/plausible signals also seemed slightly shorter in length than those considered not true (median = 10 days, range: 3–14 vs. median = 12, range: 4–14).

The longest and largest signal in terms of number of cases and geography was for lesions in Massachusetts in the spring of 2007. All 24 cases sampled for chart review consisted of uncomplicated tick bites (none qualified for a diagnosis of Lyme disease). It could not be determined whether the signal was caused by an actual increase in tick bites in the affected area, an increase in concern about ticks or Lyme disease that influenced people's health-seeking behavior there, or simply chance, although the health department reported having received several press requests about Lyme disease in early to mid-May, the period of the signal. Of possible relevance is the fact that the number of confirmed reported cases of Lyme disease in the state increased by 35% (n=875) between 2006 and 2007; perhaps tick populations and/or public awareness of the dangers of deer-tick bites grew.

Excluding the four non-GI signals in Minnesota, 19.2 hours of health-department person-time were spent investigating 13 signals—2.2 hours for three signals in California, 6.5 hours for six signals in Texas, 8 hours for one signal in Minnesota, and 2.5 hours for three signals in Massachusetts. The median investigation time per signal was 50 minutes. The eight-hour investigation was conducted by the Minnesota Department of Health, which attempted to contact by phone all 36 patients in that lower-GI signal, in addition to checking for similarities between those cases and cases identified in other outbreak investigations.

Using traditional surveillance methods in the four regions, there were 117 eligible GI illness clusters (or outbreaks) and seven eligible respiratory clusters (or outbreaks), distributed as follows: four GI clusters with eight to 43 ill in San Mateo County; five GI clusters with four to 75 ill and three respiratory clusters with two to seven ill in Austin/Travis County; 44 GI clusters with two to 17 ill and four respiratory clusters with two to 12 ill in greater Boston; and 64 GI clusters with two to 66 ill in greater Minneapolis–St. Paul. The median number ill per cluster was three for respiratory clusters, most of which consisted of just a few epidemiologically linked cases of pertussis, seven for GI clusters, and six for all clusters.

Of the 13 signals of which health departments were alerted, none was related to any of the eligible outbreaks or clusters reported and none was deemed of public health concern, notwithstanding the two true varicella clusters.

Given the RI≥365-day alerting criterion, there should by chance alone be one false alert per syndrome

**Table 2. Descriptive statistics, findings, and disposition of signals detected during an evaluation of a syndromic surveillance system operating in four U.S. metropolitan areas, 2007–2008[a]**

| State (west to east) | Syndrome | Number of observed cases | Number of expected cases | Number of zip codes | Number of days | Recurrence interval (in days) | Number of alerts per signal[b] | Time spent (in hours) | Findings | Disposition |
|---|---|---|---|---|---|---|---|---|---|---|
| California | GI, lower | 37 | 17.32 | 2 | 14 | 667 | 1 | 1 | 95% of cases in one zip code area; 80% of cases in adults; 62% of diagnoses viral enteritis or gastroenteritis/colitis, other 38% diarrhea. County health department received information late so did not investigate except to obtain line list. | Plausible, not investigated |
| California | Hemorrhagic | 10 | 2.34 | 9 | 12 | 667 | 1 | 0.3 | Symptoms included thrombocytopenia, nonthrombocytopenic purpura, nosebleed, hemorrhagic unspecified. No relationship among cases noted. | Not a true cluster |
| California | Rash | 5 | 0.54 | 2 | 14 | 10,000 | 2 | 0.8 | VAR in 6- to 12-year-olds; four of the five cases had documentation of previous vaccination; two cases in same family; at least one other case in contact with them. County health department not interested, as it tracked only those VAR cases requiring hospitalization. | True cluster |
| Texas | GI, upper | 8 | 0.98 | 14 | 6 | 455 | 1 | 3.5 | No relationship with concurrent foodborne illness outbreak. Cases involved GI procedures, medication reactions, or onsets two to eight weeks prior to visit. (Chart review done.) | Not a true cluster |
| Texas | ILI in 5- to 12-year-olds | 3 | 0.05 | 3 | 13 | 10,000 | 3 | 0 | Visits due to ILI were steadily increasing in sentinel ILI surveillance data. Three cases in three zip code areas. Investigation not deemed to be warranted. | Plausible, not investigated |
| Texas | ILI in >12-year-olds | 17 | 4.83 | 1 | 3 | 2,000 | 2 | 0 | Influenza activity in Texas during week of alert was widespread. Sentinel ILI surveillance for the week showed highest percentage of visits due to ILI since season started. Alert deemed in line with current activity. | Plausible, not investigated |
| Texas | Neurologic | 8 | 0.89 | 32 | 13 | 1,667 | 6 | 1 | Most cases aged ≥45 years; health department concluded not *Naegleria* infection, of which there had been two well-publicized cases in children. Symptoms included convulsions, diplopia, dysphagia, and alteration of awareness. Cases apparently unrelated. | Not a true cluster |

**Table 2 (continued). Descriptive statistics, findings, and disposition of signals detected during an evaluation of a syndromic surveillance system operating in four U.S. metropolitan areas, 2007–2008[a]**

| State (west to east) | Syndrome | Number of observed cases | Number of expected cases | Number of zip codes | Number of days | Recurrence interval (in days) | Number of alerts/signal[b] | Time spent (in hours) | Findings | Disposition |
|---|---|---|---|---|---|---|---|---|---|---|
| Texas | Respiratory, 1 | 16 | 3.56 | 4 | 3 | 1,667 | 2 | 0 | Not investigated due to insufficient person-time. | Not investigated |
| Texas | Respiratory, 2 | 18 | 4.53 | 1 | 6 | 1,667 | 3 | 2 | Activity seemed normal for winter respiratory season. Nothing unusual noted in line list. | Plausible, not investigated |
| Minnesota | GI, lower | 36 | 14.39 | 7 | 12 | 1,667 | 2 | 8 | Health department attempted interviews with all, interviewed eight, of which several were children. A couple of restaurants were reported but none that were in common. Cross-referencing of all information reported with that from other current investigations produced no matches. | Not a true cluster |
| Minnesota | Hemorrhagic, 1 | 5 | 0.38 | 9 | 4 | 625 | 3 | 0[c] | Only one was case of unexplained bleeding. Cases unrelated. (Chart review done.) | Not a true cluster |
| Minnesota | Hemorrhagic, 2 | 4 | 0.13 | 20 | 3 | 5,000 | 8 | 0 | Four cases in a 20-zip-code area. Based on experience with above alert, not considered worth investigating. | Not investigated |
| Minnesota | ILI in <5-year-olds | 3 | 0.04 | 13 | 4 | 1,111 | 2 | 0 | Three cases in a 13-zip-code area in October. Not considered worth investigating. | Not investigated |
| Minnesota | ILI in 5- to 12-year-olds | 3 | 0.09 | 1 | 14 | 385 | 1 | 0[c] | Occurred in late August/early September. Symptoms included pneumonia, pharyngitis, and strep throat. Cases apparently unrelated. (Chart review done.) | Not a true cluster |
| Massachusetts | GI, upper | 32 | 9.85 | 14 | 7 | 10,000 | 7 | 0 | Occurred at end of community-wide norovirus outbreak. Not investigated. | Plausible, not investigated |
| Massachusetts | Lesions | 106 | 61.14 | 88 | 14 | 10,000 | 16 | 2 | Occurred in spring. All coded as uninfected insect bite. All reviewed cases had tick bite, no diagnoses of tick-borne illness. (Chart review done for 24.) | Plausible |
| Massachusetts | Rash | 4 | 0.15 | 2 | 6 | 1,667 | 2 | 0.5 | Three cases of breakthrough VAR, including two in same household, not deemed of public health importance; fourth case was documentation of VAR immunity, not acute disease. (Chart review done.) | True cluster |

[a]Surveillance periods: California—January 22, 2007, to January 21, 2008; Texas and Minnesota—February 8, 2007, to February 7, 2008; Massachusetts—March 27 to July 26, 2007.

[b]In the case of alerts in the same syndrome that overlap in space and time, the details listed were for the alert with the highest recurrence interval and, if there were ties, the highest number of cases and, if there were ties, the earliest end date.

[c]The investigation was carried out by the clinical responder at the health-care organization, not the health department, so it is not reflected in this Table.

GI = gastrointestinal

VAR = varicella

ILI = influenza-like illness

per year for each of the five organizations. Considering the number of syndromes and the length of surveillance, this translates to an estimated 36 false alerts by chance for the whole system (not all syndromes were tracked for all five data-providing organizations), compared with 62 alerts observed. If we had instead used an alerting threshold of RI≥10,000 days, 1.3 false alerts would have been expected, compared with 15 alerts observed. These 15 alerts belonged to four discrete signals, of which one was judged true and three were judged plausible outbreaks. One of these three was the Massachusetts tick-bite cluster, which contributed 10 of the 15 alerts with RI≥10,000 days. The other two plausible signals were assumed by health department staff to be related to ILI and norovirus known to be circulating in the respective communities.

## DISCUSSION

This ambulatory care-based syndromic surveillance system did not identify any disease clusters of interest to the corresponding public health agencies during up to one year of real-time operation in four geographically disparate metropolitan areas. Why did we not detect any of the clusters found through traditional surveillance, given that there were dozens in two of the surveillance areas and that the system was capable of detecting at least two clusters of true disease? A number of factors may have been involved:

- **Patterns in infectious disease occurrence during the surveillance period:** The fewer or smaller the outbreaks and/or the less severe the disease, the lower the chances of detecting an outbreak. For example, San Mateo County had only four noninstitutional outbreaks/clusters during the surveillance year, all due to either confirmed or probable norovirus, for which many of the affected typically do not seek medical care.[14] The largest of these was in the setting of a children's center, with 43 ill, and it was noted that the duration of symptoms was short and there were no hospitalizations. The median outbreak/cluster size for all jurisdictions combined was only six.

- **Population coverage of the participating health-care organizations:** In Minnesota, at least some of the reported eligible outbreaks/clusters of GI illness were not tiny, were localized in time and space, and likely or certainly led to patients seeking care at health-care organizations, all of which would have favored their detection. They included an outbreak of *Escherichia coli* O157:H7 from a pig roast, with 26 ill; *Shigella sonnei* in two day care centers, with about 20 medically

attended cases; and an outbreak of *Salmonella* Muenchen of unknown origin, with 20 ill. However, HealthPartners covered only about 10% of the population in the Minneapolis–St. Paul area. If 10% of the 20 medically attended cases of *Shigella sonnei* in the day care centers sought care at HealthPartners, it would have been very difficult to detect this as a statistically significant increase in a syndrome as common as lower GI illness, even if the two cases had been in the same zip code.

Unpublished simulation work using data from this surveillance system from previous years has shown very modest sensitivity to detect even large outbreaks (Unpublished report. Kleinman KP, Abrams AM, Yih WK, Heisey-Grove D, Matyas BT, Nordin JD, et al. A simulation experiment of syndromic surveillance of outpatient visits for public health surveillance of non-catastrophic events using multiple scenarios and replicated in four systems. 2008). It stands to reason that better coverage would improve sensitivity, but at what level of coverage sensitivity would improve remains to be demonstrated.

- **Type of spatial data captured:** Zip code of residence may not be the most relevant spatial data to use, especially in urban areas where people often are exposed at or near their workplace, school, or day care center.

- **Type of medical encounter captured:** Even where disease outbreaks were frequent, disease was severe enough to cause people to seek care, and participating health-care organizations had high population coverage, the ability of this system to detect outbreaks would depend on whether patients tended to seek care in ambulatory clinics as opposed to EDs. If some health-seeking patients chose the ED, the ability to detect a cluster in ambulatory care data would be diminished.

- **Syndrome definitions:** Broad syndrome definitions, which were intended to maximize sensitivity at the level of individual cases at the expense of specificity, can potentially decrease the signal-to-noise ratio and obscure signals.[23] Colleagues at the Massachusetts Department of Public Health scrutinized the definitions the system had been using since its inception and eliminated some codes. These more specific definitions were employed for all sites during the evaluation period. However, some syndromes were still quite broad (e.g., respiratory, lower GI) and, depending on one's goal, could have been made

more specific to create, for example, a bloody diarrhea syndrome instead of lower GI.

- **Inconsistencies in diagnostic coding:** Mistakes or inconsistencies in diagnostic coding can lead to either the creation of false signals or the dissipation of would-be real signals. An example of the former was the generation of false alarms in a hemorrhagic syndrome in the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) due to the diagnostic miscoding of patients undergoing anticoagulation therapy as having "coagulation defect, not otherwise specified."[24] An example of the latter appears in an evaluation by the New York City Department of Health and Mental Hygiene, in which 15 patients were transferred to a participating ED on one day during a GI illness outbreak (ultimately confirmed as due to norovirus) at their nursing home, and yet no signal was generated. Field investigation found that the ED had recorded visit dates spanning three days instead of one and had mistakenly entered the nonspecific (and non-GI-related) emergency medical service call type instead of chief complaint, which on chart review was found to be consistent with gastroenteritis for most of the patients.[25] We do not know the extent to which miscoding caused false signals or suppressed real ones in our system.

- **Aberration detection method:** It is of course also possible that other cluster detection methods, whether spatio-temporal or purely temporal, might have performed better than the space-time permutation method used in this study.

On the positive side, two rash signals in two states were found to be associated with linked varicella cases—evidence that the system is capable of detecting at least certain types of disease clusters or outbreaks. The number of expected false alarms at the selected threshold of RI≥365 days was 36, compared with 62 total alerts grouped into 17 signals, which corresponded to a mean of 0.47 signals per syndrome-jurisdiction during a one-year period. The 15 alerts that would have been seen had the highest threshold (RI≥10,000 days) been used were clumped into one true and three plausible signals of clusters or outbreaks.

## CONCLUSIONS

In this collaboration among academia, health-care organizations, and state and local health departments, we evaluated the performance of an ambulatory care-based real-time syndromic surveillance system in four geographically distant U.S. metropolitan areas over the course of a year. None of the 17 signals corresponded to clusters detected through traditional surveillance, nor did any lead to the discovery of new outbreaks of interest for public health. Of the 13 signals of which health departments were notified, the median investigation time per signal on the part of health department staff was 50 minutes. Possible reasons that none of the otherwise known clusters was detected include contextual factors, such as lack of large or severe enough outbreaks during the surveillance period, as well as features of the surveillance system, such as limited population coverage and the particular syndrome definitions used. However, the system demonstrated its ability to detect real clusters of illness and did not generate an unreasonable number of false alarms. Its utility for early detection of large outbreaks, such as the Milwaukee cryptosporidiosis outbreak of 1993, is still untested. The application of these methods to electronic laboratory data, which are centralized and quite complete in many states, may also be a fruitful avenue to explore, moving from syndromic to pathogen-specific surveillance.

## REFERENCES

1. Sosin DM. Draft framework for evaluating syndromic surveillance systems. J Urban Health 2003;80(2 Suppl 1):i8-13.
2. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep 2004;53(RR05):1-11.
3. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1:77-81.
4. Buehler JW, Sonricker A, Paladini M, Soper P, Mostashari F. Syndromic surveillance practice in the United States: findings from a survey of state, territorial, and selected local health departments. Adv Dis Surveill 2008;6:1-20. Also available from: URL: http://www

.syndromic.org/ADS/prepub/Buehler20080503.pdf [cited 2009 Feb 25].

5. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City [published erratum appears in Emerg Infect Dis 2006;12:1472]. Emerg Infect Dis 2004;10:858-64.

6. Balter S, Weiss D, Hanson H, Reddy V, Das D, Heffernan R. Three years of emergency department gastrointestinal syndromic surveillance in New York City: what have we found? MMWR Morb Mortal Wkly Rep 2005;54(Suppl):175-80.

7. Terry W, Ostrowsky B, Huang A. Should we be worried? Investigation of signals generated by an electronic syndromic surveillance system—Westchester County, New York. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):190-5.

8. Sheline KD. Evaluation of the Michigan Emergency Department Syndromic Surveillance System. Adv Dis Surveill 2007;4:265.

9. 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. Adv Dis Surveill 2006;1:75.

10. Costa MA, Kulldorff M, Kleinman K, Yih WK, Platt R, Brand R, et al. Comparing the utility of ambulatory care and emergency room data for disease outbreak detection. Adv Dis Surveill 2007;4:243.

11. Yih WK, Caldwell B, Harmon R, Kleinman K, Lazarus R, Nelson A, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):43-9.

12. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Public Health 2006;6:235.

13. Daniel JB, Heisey-Grove D, Gadam P, Yih W, Mandl K, DeMaria A Jr, et al. Connecting health departments and providers: syndromic surveillance's last mile. MMWR Morb Mortal Wkly Rep 2005;54(Suppl):147-50.

14. 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.

15. Ritzwoller DP, Kleinman K, Palen T, Abrams A, Kaferly J, Yih W, et al. Comparison of syndromic surveillance and a sentinel provider system in detecting an influenza outbreak—Denver, Colorado, 2003. MMWR Morb Mortal Wkly Rep 2005;54(Suppl):151-6.

16. Centers for Disease Control and Prevention (US). Syndrome definitions for diseases associated with critical bioterrorism-associated agents [cited 2008 Oct 21]. Available from: URL: http://www.bt.cdc.gov/surveillance/syndromedef/index.asp

17. Brammer TL, Murray EL, Fukuda K, Hall HE, Klimov A, Cox NJ. Surveillance for influenza—United States, 1997–98, 1998–99, and 1999–00 seasons. MMWR Surveill Summ 2002;51(SS07):1-10.

18. Platt R, Bocchino C, Caldwell B, Harmon R, Kleinman K, Lazarus R, et al. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. J Urban Health 2003;80(Suppl 1):i25-31.

19. Platt R. Homeland security: disease surveillance systems. Testimony submitted to the U.S. House of Representatives Select Committee on Homeland Security. 2003 Sep 24 [cited 2008 Oct 21]. Available from: URL: https://btsurveillance.org/btpublic/Publications/house_testimony.pdf

20. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005;2:e59.

21. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. Am J Epidemiol 2004;159:217-24.

22. SaTScan. SaTScan™: Version 7.0 [cited 2008 Oct 21]. Available from: URL: http://www.satscan.org

23. Pendarvis J, Miramontes R, Schlegelmilch J, Fleischauer AT, Gunn J, Hutwagner L, et al. Analysis of syndrome definitions for gastrointestinal illness with ICD-9 codes for gastroenteritis during the 2006–07 norovirus season in Boston. Adv Dis Surveill 2007;4:263.

24. Oda G, Hung L, Lombardo J, Wojcik R, Coberly J, Sniegoski C, et al. Diagnosis coding anomalies resulting in hemorrhagic illness alerts in Veterans Health Administration outpatient clinics. Adv Dis Surveill 2007;4:260.

25. Steiner-Sichel L, Greenko J, Heffernan R, Layton M, Weiss D. Field investigations of emergency department syndromic surveillance signals—New York City. MMWR Morb Mortal Wkly Rep 2004;53(Suppl):184-9.