Available online at www.sciencedirect.com
ScienceDirect
Journal of Biomedical Informatics 40 (2007) 370–379
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Methodological Review

Outbreak detection through automated surveillance: A review of the determinants of detection

David L. Buckeridge *

McGill Clinical and Health Informatics, Department of Epidemiology and Biostatistics, McGill University, 1140 Pine Avenue West, Montreal, Que., Canada H3A 1A3

Received 28 April 2006
Available online 5 October 2006

Abstract

Public health agencies and other groups have invested considerable resources in automated surveillance systems over the last decade. These systems generally follow syndromes in pre-diagnostic data drawn from sources such as emergency department visits. A main goal of syndromic surveillance systems is to detect outbreaks rapidly and the number of studies evaluating outbreak detection has increased recently. This paper reviews these studies with the goal of identifying the determinants of outbreak detection in automated syndromic surveillance systems. The review identified 35 studies with 22 studies (63%) relying on naturally occurring outbreaks and 13 studies (37%) relying on simulated outbreaks. In general, the results from these studies suggest that syndromic surveillance systems are capable of detecting some types of disease outbreaks rapidly with high sensitivity. The determinants of detection included characteristics of the system and of the outbreak. Influential system characteristics included representativeness, the outbreak detection algorithm, and the specificity of the algorithm. Important outbreak characteristics included the magnitude and shape of the signal and the timing of the outbreak. Future evaluations should aim to address inconsistencies in the evidence noted in this review and to identify the potential influence of other factors on outbreak detection.

Keywords: Surveillance; Evaluation; Outbreak detection; Medical informatics; Public health

1. Introduction

Over the last decade, public health agencies and others have invested substantial resources in the development and operation of automated systems for syndromic surveillance [1–5]. These systems monitor pre-diagnostic data drawn from sources such as emergency department visits and sales of pharmaceuticals. While syndromic surveillance systems may be effective for many purposes, a main purpose of these systems is usually to detect disease outbreaks rapidly.

Given that outbreak detection is a fundamental reason for conducting syndromic surveillance, public health personnel need to understand how well syndromic surveillance systems accomplish this goal. The relevant question from a public health perspective is not simply whether syndromic surveillance can detect an outbreak, but rather what are the determinants of outbreak detection when using syndromic surveillance. In other words the people who must decide whether to install a syndromic surveillance system and those who must operate the systems need to know the most influential factors of system performance.

A recent systematic review noted few published evaluations of syndromic surveillance systems [6]. More recently, however, the evaluation of outbreak detection through syndromic surveillance has received considerable attention. A working group established by the CDC released evaluation guidelines for these systems recently [7] and the number of published studies of outbreak evaluation has increased over the last few years.

The goal of this paper is to review studies that have evaluated outbreak detection through automated syndromic surveillance. The specific aims are to summarize the

* Corresponding author. Fax: +1 514 843 1551.
E-mail address: david.buckeridge@mcgill.ca

1532-0464/$ - see front matter
doi:10.1016/j.jbi.2006.09.003

evidence relating to the determinants of outbreak detection and to examine the merits of different methods for evaluating outbreak detection. Public health surveillance is a process that includes many actions other than analysis to detect outbreaks, including identifying cases, gathering data, and guiding decisions about public health actions. While it is never possible to isolate fully one component of the overall surveillance process from the other components, the focus of this review is to synthesize research findings related to outbreak detection. The paper is intended to inform further research on outbreak detection and to provide public health practitioners with a summary of evidence for one aspect of syndromic surveillance.

2. Methods

The scope of the review included studies that evaluated the use of routinely collected, pre-diagnostic data for outbreak detection. To be included in the review, a study had to apply an outbreak detection algorithm prospectively to syndromic data. In other words, the evaluator had to make an alarm decision at a given time using only data collected prior to that time. The rationale for this inclusion criterion was to ensure that the approach to surveillance evaluated in a study could be reproduced in public health practice. Studies were excluded, therefore, if they used data that were diagnostic, or data that were acquired through active or manual collection. In addition, studies were required (1) to compare syndromic data to some 'gold standard' data with known outbreaks, or (2) to inject simulated outbreaks into the syndromic data. Any system description without evaluation was therefore excluded, as were studies that relied on wholly simulated data, and studies that did not apply an outbreak detection algorithm. There had to be a reproducible decision rule for declaring an outbreak. These criteria excluded studies that compared syndromic data to other data sources through cross-correlation analysis. Finally, to ensure that the evaluated data were still likely to be available in practice, only studies published within that last 20 years were considered.

An initial set of studies was identified through a recent systematic review of syndromic surveillance systems [6]. For each of the studies in the systematic review, PubMed was searched using the *Related Articles* feature and the 100 articles most similar to each study in the systematic review were considered for inclusion in this review. The *Related Article* feature uses a variant of the cosine similarity method to identify other articles with similar terms in the title, abstract or keywords [8]. PubMed was also used to search for eligible studies using the query "'syndromic surveillance' AND (evaluation OR test)." Literature searches were conducted in April of 2006. Online resources were also consulted, including the www.syndromic.org Internet site, which is maintained by the International Society for Disease Surveillance, and the annotated bibliography published by the Centers for Disease Control and Prevention (http://www.cdc.gov/epo/dphsi/syndromic/index.htm). All papers published in the proceedings of the syndromic surveillance conferences between 2001 and 2004 were also reviewed manually for eligibility. Finally, the references in all studies included in this review were examined and considered for inclusion.

3. Results

3.1. Descriptions of studies

The search identified 35 studies described in 34 published articles [9–42]. Table 1 summarizes the studies in terms of their general characteristics, which are presented below.

3.1.1. Study setting

Studies tended to be set in metropolitan areas in the United States, but some were from France, England and Canada. There was a trend towards increasing numbers of studies over time with the majority of studies published in 2004 or later. Records from emergency department visits were analyzed most commonly (16 studies, 46%), but data from OTC sales (8 studies, 23%) and ambulatory visits (8 studies, 23%) were also used in some studies. Data were frequently grouped into a syndrome that reflected a respiratory (19 studies, 54%) or gastrointestinal illness (12 studies, 34%), with most of the studies of gastrointestinal illness (12 of 13 studies) relying on naturally occurring outbreaks for evaluation. There was no consistent approach across studies to grouping records into syndromes.

3.1.2. Outbreak signals in test data

Most studies (22 studies, 63%) used naturally occurring outbreak signals to evaluate outbreak detection, while a smaller number relied on simulated signals (13 studies, 37%). Both types of studies tended to rely on a small amount of authentic data for the evaluation (Fig. 1), with most using under 2 years of data. In studies where authors relied on naturally occurring outbreaks the tendency was to evaluate detection of annually occurring respiratory or gastrointestinal outbreaks. These outbreaks appear in syndromic surveillance time series as a gradual increase in incidence over weeks or months with a peak incidence many times larger than the baseline incidence at other times of the year (Fig. 2). In studies where authors used injected signals, the tendency was to model the signal after an anthrax outbreak or a mathematical function, such as a step-function or an exponential distribution (Table 2). The methods used to simulate signals ranged from simple mathematical functions to more complex stochastic simulation models with more recent studies tending to rely on more detailed models.

3.1.3. Outbreak detection algorithms

Most of the studies (24, 69%) used temporal detection algorithms only, while the remaining studies used spatial or space–time detection algorithms (Table 3). All studies

Table 1
Summary of studies

Author	Year	Syndromic data	Location	Syndrome
Naturally occurring outbreaks				
Marx et al. [23]	2006	ED visits, Absenteeism, OTC	NYC	GI
Kulldorf et al. [22]	2005	ED visits	NYC	Diarrhea
Das et al. [12]	2005	OTC	NYC	ILI, GI
Chen et al. [11]	2005	Prescription pharmaceuticals	NY State	Macrolides
Ritzwoller et al. [34]	2005	Ambulatory visits	Denver, CO	ILI
Yih et al. [42]	2005	Ambulatory visits	Minneapolis, St. Paul, MN	GI
Balter et al. [9]	2005	ED visits	NYC	GI
Fleming et al. [14]	2004	Ambulatory visits	England, UK	Asthma
Heffernan et al. [17]	2004	ED visits	NYC	Respiratory, Fever, GI
Siegrist and Pavlin [36]	2004	Ambulatory visits, OTC, Prescription pharmaceuticals	USA	Respiratory, GI
Edge et al. [13]	2004	OTC	SK, ON	Gastrointestinal
Hogan et al. [18]	2003	OTC	PA, IN, UT	Electrolytes
Irvin et al. [19]	2003	ED visits	Detroit, MI	Anthrax
Ivanov et al. [20]	2003	ED visits	Utah	Respiratory, GI
Mostashari et al. [25]	2003	Ambulance dispatches	NYC	ILI
Mostashari et al. [26]	2003	Dead bird reports	NYC	Death
Weber and Pitrak [41]	2003	ED visits	Chicago, IL	Potential WNV
Goldenberg et al. [15]	2002	OTC	Pittsburgh, PA	Cough and cold
Harcourt et al. [16]	2001	Nurse hotline calls	England, UK	Respiratory, ILI
Tsui et al. [38]	2001	ED visits	Pittsburgh, PA	Respiratory, ILI
Quenel and Dab [29]	1998	Prescription pharmaceuticals	Paris, France	Respiratory
Rodman et al. [35]	1998	Nurse hotline calls	Milwaukee, WI	Diarrhea
Simulated outbreaks				
Wang et al. [40]	2005	ED visits	Boston, MA	Respiratory
Wallstrom et al. [39]	2005	OTC	Washington, DC	Gastrointestinal
Kleinman et al. [21]	2005	Ambulatory visits, ED visits	Boston, MA	Respiratory
Buckeridge et al. [10]	2005	Ambulatory visits	Norfolk, VA	Respiratory
Nordin et al. [27]	2005	Ambulatory visits	Minneapolis, St. Paul, MN	Respiratory
Miller et al. [24]	2004	Ambulatory visits	Minneapolis, St. Paul, MN	ILI
Ozonoff et al. [28]	2004	ED visits	Boston, MA	URI
Reis and Mandl [31]	2004	ED visits	Boston, MA	Respiratory
Stoto et al. [37]	2004	ED visits	Washington, DC	ILI
Reis and Mandl [30]	2003	ED visits	Boston, MA	Respiratory
Reis and Mandl [32]	2003	ED visits	Boston, MA	All visits
Reis et al. [33]	2003	ED visits	Boston, MA	All visits
Goldenberg et al. [15]	2002	OTC	Pittsburgh, PA	Cough and cold

ED, emergency department; OTC, over-the-counter pharmaceuticals; GI, gastrointestinal; ILI, influenza-like illness; URI, upper respiratory tract infection; and WNV, West Nile virus.

evaluated surveillance of a single data source, except for one study that considered concurrent surveillance of three data sources [36]. The majority of studies applied a decision rule daily (28 studies, 80%), although some analyzed the data weekly (6 studies, 17%) or monthly (1 study, 3%). Most studies conducted a temporal analysis using a regression method that modeled daily counts for a single surveillance data source. Statistical process control methods were used more frequently for analyses of weekly or monthly counts. Methods used for spatial and space–time analyses were varied, but a scan statistic was used most frequently.

3.2. Evaluation of outbreak detection using natural outbreaks

Most studies that relied on natural outbreaks reported that surveillance using syndromic data tended to detect the signals seen in the comparison data, often immediately before, or near the beginning of the signal detected in the gold standard data source. For example, one author reported an alarm from surveillance of outpatient clinic visits for asthma in the same week as an unusual surge in pollen counts [14]. Another author reported that an alarm from surveillance of the incidence of calls to a nurse hotline for influenza-like illness occurred in the same week as an increase in influenza isolates in the community [16].

Five studies examined detection performance on a number of outbreaks sufficiently large to report quantitative measures of sensitivity and timeliness [9,18,20,26,36]. The authors of a study of over-the-counter electrolyte sales reported that weekly temporal surveillance of these data had sensitivity and specificity of 100% for the three seasonal outbreaks at five geographic locations, as compared to surveillance of pediatric hospital admissions for outbreaks of gastrointestinal infections [18]. This finding is difficult to generalize to detection of smaller, sporadic outbreaks because the authors set the decision threshold for the Exponentially Weighted Moving Average (EWMA) method to 9 standard deviations above the mean, and an algorithm with

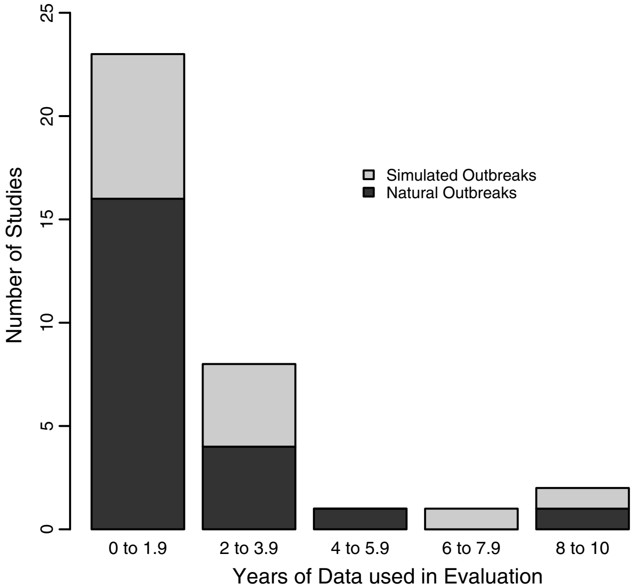

Fig. 1. Number of years of baseline data used in evaluation studies.

such a high decision threshold may not detect smaller outbreaks. In a similar study, the same authors compared free-text chief complaints from ED visits to hospital admissions for surveillance of respiratory and gastrointestinal outbreaks [20]. They reported that using a EWMA algorithm with a threshold 8 standard deviation above the mean detected all 3 gastrointestinal outbreaks with no false alarms. For the 3 respiratory outbreaks, the authors required a threshold of 13 standard deviations above the mean to detect all outbreaks with perfect specificity.

Another study examined simultaneous surveillance of three data sources for respiratory and gastrointestinal outbreaks with the gold-standard determined through expert review of the data [36]. The authors reported that the median time to alarm for the best detection algorithms was the first day of the outbreak with a specificity of 98% (1 false alarm every 6 weeks) and a sensitivity per outbreak of 100% (8 of 8) for respiratory outbreaks and 88% (6 of 7) for gastrointestinal outbreaks. The authors of this study also compared the performance of different outbreak detection algorithms and they noted considerable differences in the accuracy and timeliness of outbreak detection between algorithms. Wavelet and EWMA-based algorithms had the best performance. Nearly all of the signals in this study were annually occurring outbreaks, with the notable exception of one gastrointestinal outbreak presenting as a sudden spike, which was missed by most algorithms.

A third study, which examined surveillance of dead birds, reported that spatial analysis by census tract identified clusters of humans with West Nile virus (WNV) an average of 12 days before the onset of human illness and 17 days before diagnosis of human illness [26]. Sensitivity

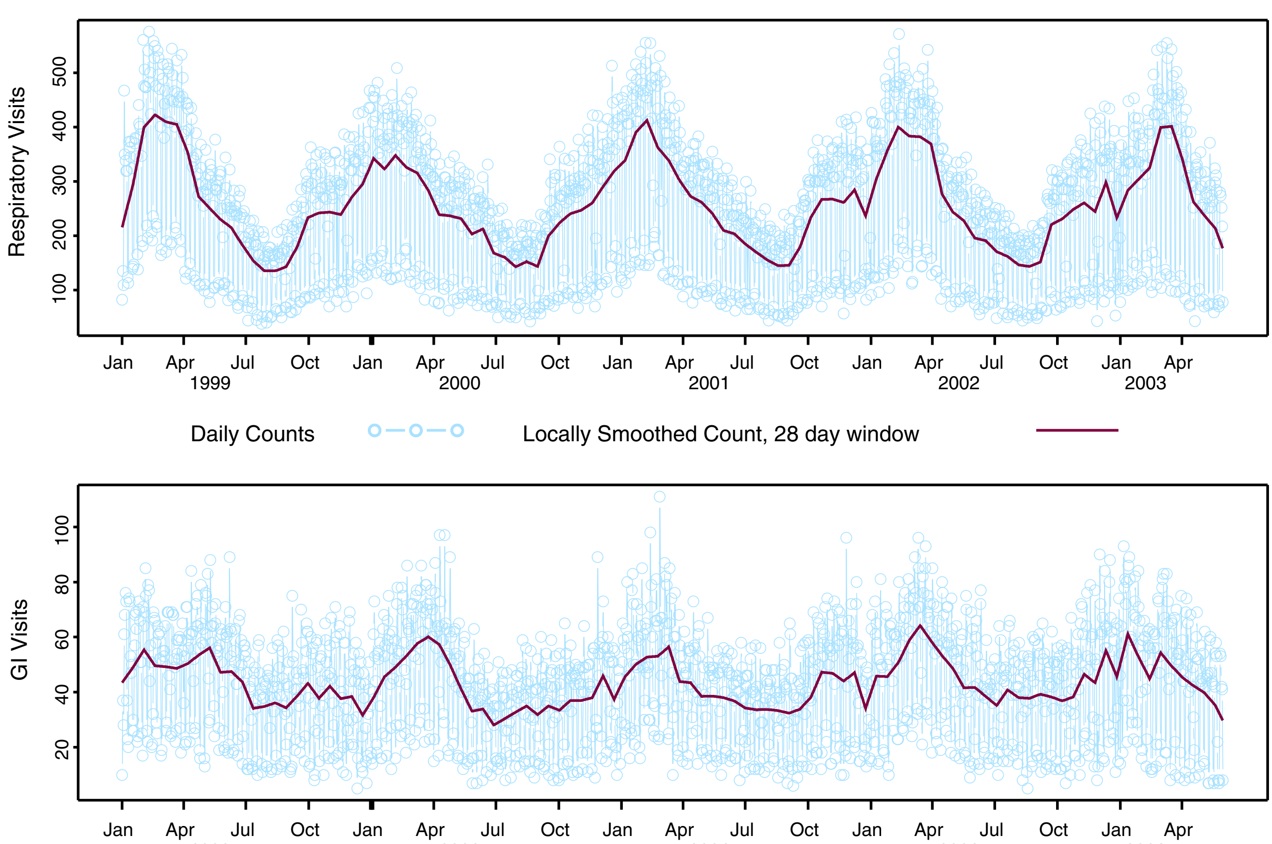

Fig. 2. Example data from ambulatory physician visits for respiratory and gastrointestinal syndromes.

Table 2
Characteristics of simulated signals used in evaluation studies

Author	Year	Signal shape	Signal duration	Signal magnitude
Signal modeled after specific disease				
Wang et al. [40]	2005	Multinomial, Linear, Exponential	7 days	5–50% of mean baseline
Wallstrom et al. [39]	2005	Historical outbreak	5 weeks	0.1–10% of population infected
Kleinman et al. [21]	2005	Model output	Various	10^{15} anthrax spores
Buckeridge et al. [10]	2005	Model output	Various	1–16% of population infected
Nordin et al. [27]	2005	Model output	Various	Various
Miller et al. [24]	2004	Historical outbreak	43 days	5–200% of mean baseline
Goldenberg et al. [15]	2002	Linear	3 days	0–200% range of baseline data
Signal modeled after mathematical function				
Ozonoff et al. [28]	2004	Step	1 day	17–34% of mean baseline
Reis and Mandl [32]	2003	Step	7 day	4–22% of mean baseline
Reis and Mandl [30,31]	2003, 2004	Step	7 days	10–50% of mean baseline
Stoto et al. [37]	2004	Linear	3, 9 days	30–300% of mean baseline
Reis et al. [33]	2003	Step, linear, exponential	3, 7, 14 days	4–33% of mean baseline

Table 3
Outbreak detection algorithms used in evaluation studies

Author	Year	Temporal	Spatial	Algorithm
Natural				
Weber and Pitrak [41]	2003	Monthly	Region	Historical mean
Fleming et al. [14]	2004	Weekly	Region	Historical mean
Edge et al. [13]	2004	Weekly	Region	SPC (MA, Cusum)
Hogan et al. [18]	2003	Weekly	Region	SPC (EWMA)
Tsui et al. [38]	2001	Weekly	Region	Regression (Serfling)
Quenel and Dab [29]	1998	Weekly	Region	Regression (SARIMA)
Harcourt et al. [16]	2001	Weekly	Sub-region (call centers)	Regression (GLM binomial)
Marx et al. [23]	2006	Daily	Region	Regression (Serfling), Scan (Temporal), SPC (Cusum)
Das et al. [12]	2005	Daily	Region	Regression (Serfllng)
Siegrist and Pavlin [36]	2004	Daily	Region	Various
Irvin et al. [19]	2003	Daily	Region	Historical mean
Ivanov et al. [20]	2003	Daily	Region	SPC (EWMA)
Mostashari et al. [25]	2003	Daily	Region	Regression, Cyclical regression at set thresholds
Goldenberg et al. [15]	2002	Daily	Region	Wavelet and AR
Rodman et al. [35]	1998	Daily	Region	Historical mean
Mostashari et al. [26]	2003	Daily	Sub-region (CT)	Scan (Spatial)
Chen et al. [11]	2005	Daily	Sub-region (County)	SPC (Cusum)
Ritzwoller et al. [34]	2005	Daily	Sub-region (ZIP)	Regression (GLMM) and Scan
Yih et al. [42]	2005	Daily	Sub-region (ZIP)	Scan (Space–Time with GLMM)
Balter et al. [9]	2005	Daily	Sub-region (ZIP)	Scan (Space–Time)
Heffernan et al. [17]	2004	Daily	Sub-region (ZIP)	Scan (Temporal and Spatial)
Kulldorf et al. [22]	2005	Daily	Sub-region (ZIP, Hospital)	Scan (Space–Time Permutation)
Injected				
Wang et al. [40]	2005	Daily	Region	Regression (ARP)
Wallstrom et al. [39]	2005	Daily	Region	Regression (ARIMA), SPC (EWMA, Cusum)
Buckeridge et al. [10]	2005	Daily	Region	Regression + SPC (ARIMA and Cusum)
Miller et al. [24]	2004	Daily	Region	Regression + SPC (AR and Cusum)
Reis and Mandl [31]	2004	Daily	Region	Regression (ARMA)
Stoto et al. [37]	2004	Daily	Region	SPC (Shewhart, EWMA, Cusum)
Reis and Mandl [30,32]	2003	Daily	Region	Regression (ARMA)
Reis et al. [33]	2003	Daily	Region	Regression + Filter (ARMA and Temporal Filters)
Goldenberg et al. [15]	2002	Daily	Region	Wavelet and AR
Ozonoff et al. [28]	2004	Daily	Sub-region (CT)	Various (ARMA, M, Bivariate)
Kleinman et al. [21]	2005	Daily	Sub-region (ZIP)	Regression (GLMM), Scan
Nordin et al. [27]	2005	Daily	Sub-region (ZIP)	Scan (Space–Time Poisson)

AR, autoregressive; ARP, autoregressive periodic; ARMA, autoregressive moving average; ARIMA, autoregressive integrated moving average; CT, census tract; EWMA, exponentially weighted moving average; GLM, generalized linear model; GLMM, generalized linear mixed model; MA, moving average; SARIMA, seasonal autoregressive integrated moving average; and SPC, statistical process control.

in this study was 82% (9 of 11 clusters) with a decision threshold of a p-value <0.10. In the final study, the authors compared gastrointestinal outbreaks detected through syndromic surveillance of ED visits to gastrointestinal outbreaks involving $\geqslant 10$ people that were identified through routine public health measures [9]. Over 3 years, none of the 49 outbreaks identified through routine measures were detected by the ED syndromic surveillance system, but some of the 236 signals from the syndromic system were attributed to true outbreaks.

Five studies that relied on naturally occurring outbreaks reported that they did not detect an outbreak identified through the gold standard comparison data. In one case, surveillance of over-the-counter cough and cold medicine sales did not detect the start of the influenza outbreak [15]. The authors hypothesized that this result may have been attributable to the influenza outbreak beginning on a holiday, and they did not report results for alarms at subsequent points in the influenza outbreak. In another study, spatial surveillance of ED visits by hospital location for diarrhea and vomiting failed to detect an institutional outbreak reported directly to public health [17]. This false negative was attributed to outbreak cases visiting a hospital that was not participating in the surveillance system, and a retrospective analysis indicated that syndromic surveillance would have detected the outbreak had that hospital been participating in the surveillance system. In a third study, spatial surveillance of dead birds aggregated to census tracts failed to identify WNV in two communities that experienced confirmed WNV in birds, mosquitoes, and humans [26]. For one of the communities, the one human case was attributed to a possible mosquito exposure outside of the community. For the second community, the one human case was a homeless person and the authors hypothesized he may have been exposed outside of the community. In the fourth study, temporal surveillance of the incidence of ED visits for conditions potentially attributable to WNV failed to detect the emergence of WNV in the Chicago area [41]. The authors do not speculate why this outbreak of 500 WNV cases occurring over three months was not detected. In the fifth study, authors sought reasons for why syndromic surveillance of ED visits did not detect any of 49 gastrointestinal disease outbreaks investigated by public health over a 3 year period [9]. In 36 outbreaks, few or no patients visited an ED and in another 2 outbreaks many of those involved were visitors who left the region before the onset of symptoms. For the remaining 11 outbreaks, patients visited EDs, but for 3 outbreaks the EDs were not part of the surveillance system. Other reasons noted for the lack of sensitivity were that outbreak cases visited EDs over a number of days or weeks, or many cases were coded using a single record.

Authors reported an empirical estimate of specificity in only one study [36]. In that study, experts identified the location of outbreaks by visual inspection of the data, and the authors reported high sensitivity and low timeliness in the detection of naturally occurring respiratory and gastrointestinal outbreaks at a specificity of 98%. Although authors of other studies did not estimate specificity empirically, some did report estimates of positive predictive value. In one study, authors reported that over 80% of alarms from citywide surveillance of diarrhea and vomiting visits to emergency departments occurred during outbreaks confirmed by laboratory testing [17]. In the same study, however, the authors reported that alarms from spatial surveillance of fever and respiratory visits to emergency departments occurred with the same frequency during and outside of influenza outbreaks. Another study reported that over 80% of alarms occurring through surveillance of ambulance dispatches for respiratory conditions occurred shortly before of during an influenza outbreak [25]. In other studies authors reported positive predictive values ranging from 0% [41] to 50% [19,38], but all of these studies relied on data with only one outbreak.

3.3. Evaluation of outbreak detection using injected signals

Authors reported sensitivity per outbreak of 100% at high specificity (97–99%) for signals with a magnitude from 20 to 300% of the average daily count in the background data. Miller [24] reported detecting all three injected lognormal signals on the day when the magnitude of the added signal was 12% of the daily background with a specificity of 98%. Another study using a 3-day step-function signal [32], reported a sensitivity of 100% at a specificity of 97% when the signal was 21% of the average daily background. Similar results were reported for a study using a 3-day linearly increasing signal, which described detecting all outbreaks with a magnitude of 36% or more of the daily background at an undefined specificity [15].

In contrast to the studies reporting high sensitivity for smaller signals, some authors reported that a signal with a larger magnitude was required to detect a large percentage of outbreaks [28,37,40]. In the study by Wang, an outbreak signal with a magnitude of 60–80% of the average daily baseline was required for a sensitivity of 100% at a specificity of 96% [40]. In the study by Stoto, a 3-day signal outside of the influenza season was not detected with certainty until it rose to 300% of the average daily background count [37]. One difference between the studies reporting high sensitivity at low signal magnitude and the studies requiring larger signals was the volume of records in the baseline data. Authors of studies that reported high sensitivity with proportionally smaller signals used baseline data with a higher daily mean, approximately 100 records each day, as opposed to mean daily counts of under 40 and as low as 3 in one study where larger signals were required to obtain high sensitivity [37].

Authors of four studies reported the time from onset of an outbreak until detection [10,21,24,27]. In all cases, the authors used a simulated anthrax outbreak. Miller reported detection of an outbreak resulting in 308 additional emergency department visits between 4 and 7 days following the release of anthrax [24]. The other authors reported earliest detection 2 days following release with the majority detected within 4 days and nearly all detected within 9 days of exposure [10,21,27].

Authors examined systematically the impact of signal magnitude on outbreak detection in nine studies [10,15,27,28,30,32,33,39,40]. In all of the studies, authors reported improved sensitivity or timeliness as the signal strength increased. For example, Ozonoff reported that doubling the signal magnitude from 17% of the mean daily baseline (35 counts) to 34% increased the sensitivity of a time-series algorithm for a 1-day signal from 13% to 29% [28]. Reis used a step-function signal of 7 days duration and reported a larger increase in sensitivity per day, from approximately 15% to nearly 70%, when increasing signal magnitude from 10 to 30% of the mean daily baseline of 60 counts [30]. Other factors that had the effect of altering the magnitude of the signal, including the proportion of ill individuals that seek care [10] and the proportion of the population covered by the system [27,39] also influenced sensitivity.

The influence of different signal shapes on outbreak detection was examined in three studies [33,37,40]. Stoto [37] injected 18 additional cases over 3 days (3, 6, and 9) or over 9 days (1, 1, 2, 2, 2, 3, 3, and 3) to represent 'fast' and 'slow' outbreaks, and found that the fast outbreaks were detected more often and more quickly than were the slow outbreaks. Wang evaluated detection of three signal shapes, a multinomial distribution patterned after a historical outbreak, a linear increase in incidence, and an exponential increase in incidence [40]. The multinomial signals were detected more quickly (70% within 3 days of release), followed by the linear signals (50% within 3 days) and the exponential signals (27% within 3 days). Reis [33] examined linear, exponential and step increases of 3, 7, and 14 days duration, and reported a difference in detection performance between the different signal shapes, but did not describe the nature of the difference.

The effect of day-of-week or season on outbreak detection was examined in five studies [24,27,28,37,40]. In general, most authors reported that for a given signal, sensitivity was higher when baseline counts were lower. Miller reported that the time to outbreak detection of the same lognormal signal in records of ED visits for respiratory conditions differed by season with detection occurring 4 days following a release in June, 5 days following a release in April, and 7 days following a release in December [24]. Stoto found that sensitivity for 'fast' signals was 100% during the summer and 20% during the winter [37]. Ozonoff reported improved sensitivity on holidays and weekends when the mean count was 15 as compared to weekdays when the mean count was 43 [28]. Nordin noted a similar relationship, with higher sensitivity in the summer and on week-ends, but observed that the influence of season and day-of-week weakened as signal strength increased [27].Wang assessed overall detection accuracy, and noted that accuracy of detection was greater in the winter (90%) than in the summer (77%) and the reason for this apparently contradictory finding is not clear [40].

The effect of different detection algorithms on sensitivity or timeliness was examined in six studies [21,28,31,33,37,39]. Two studies led by the same author examined the influence of different temporal filters applied to forecast residuals from a time-series model [31,33]. To detect a step signal of 7-day duration at a set specificity, a 7-day linear filter applied to regression residuals enhanced sensitivity over a 1-day filter (an increase in sensitivity from 30 to 71% for a signal 15% over the mean daily baseline), and tended to outperform 7-day moving average and exponential filters [33]. In another study, authors examined four statistical process control (SPC) methods and reported that smoothing algorithms, such as an exponentially weighted moving average or a cumulative sum, tended to outperform one-day threshold algorithms, especially for slowly increasing signals [37]. In a study where both temporal and spatial detection algorithms were compared, the authors demonstrated that a spatial detection algorithm, the inter-point distance M statistic, had higher sensitivity than a seasonal time-series model, when clusters were concentrated in space [28]. For more diffuse clusters, the temporal statistic tended to have higher sensitivity. In a study that compared two space–time algorithms, a model-based scan statistics was found to outperform a regression algorithm [21].

4. Discussion

This review identified 35 evaluations of outbreak detection in automated syndromic surveillance systems. The results from studies that relied on naturally occurring outbreaks suggest that automated syndromic surveillance can detect large seasonally occurring outbreaks with sensitivity and timeliness comparable to or better than systems that rely on diagnostic data. Results from studies that relied on simulated outbreak signals also suggest that syndromic surveillance systems can detect some types of disease outbreaks rapidly with high sensitivity and specificity.

The results from the different studies are useful for identifying some of the determinants of outbreak detection. In other words, the results offer evidence to support which factors will determine if and when a syndromic surveillance system will detect an outbreak. To help place the results from the current study in context, it is helpful to consider them within a broader conceptual framework for the determinants of outbreak detection. Fig. 3 presents such a framework that summarizes potential determinants of outbreak detection [7,43–45].

Factors related to baseline and outbreak cases are fundamental determinants of detection. The incidence and variation over time, space and other attributes, of baseline cases (i.e., cases not attributable to an outbreak) will influence outbreak detection. In addition, the incidence, variation and the timing of onset of outbreak cases, and the relationship between baseline and outbreak factors will also influence outbreak detection. For outbreak detection, the baseline cases can be thought of as the 'noise' and the

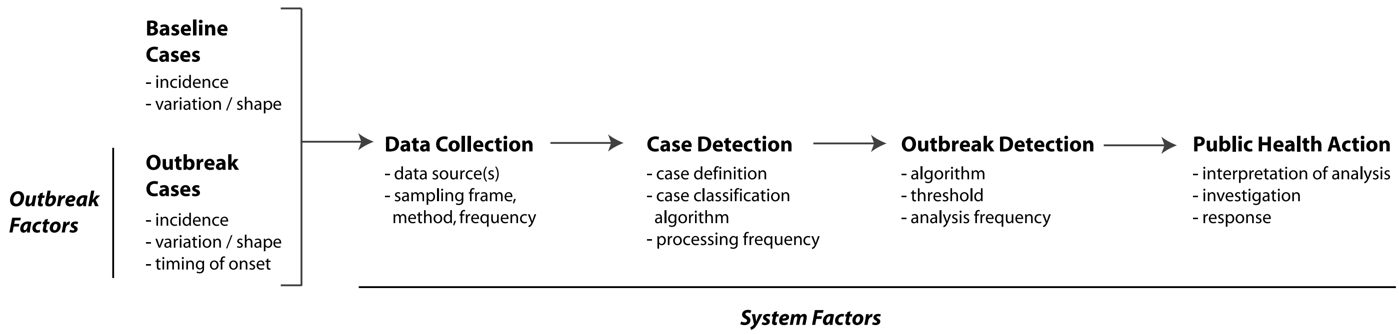

Fig. 3. A conceptual framework for the determinants of outbreak detection through automated surveillance [7,43–45]. The arrows indicate the flow of information in the surveillance process.

outbreak cases, the 'signal'. The goal of outbreak detection is to identify a signal rapidly with high accuracy when the true distinction between baseline and outbreak cases is not known. In general terms, outbreaks will be easier to detect when the incidence and variation of baseline cases are low relative to the outbreak cases.

Another set of factors relate to the surveillance system itself. The data source or sources used in a system are important for their conceptual relationship to an incident case of disease and the often implicit sampling considerations associated with a data source, including the sampling frame and frequency. Data sources that relate closely to the incidence of disease and that are collected frequently for a large proportion of the population are likely to facilitate outbreak detection. The approach taken to case detection, including the case definition and the algorithm used to infer if an individual meets a case definition are also potential determinants of outbreak detection. Once cases are identified, the temporal or space–time algorithm used to analyze the data and the threshold at which the algorithm operates may also influence outbreak detection. Finally, the public health response to analysis results, including investigation and response protocols will help to determine if and when outbreaks are detected.

The determinants of detection noted in this review include attributes of the surveillance system and attributes of the outbreak. Characteristics of the system that influenced detection included the choice of data source, the representativeness or sampling strategy of the system, the detection algorithm, and the specificity or threshold at which an algorithm operated. Authors of studies using natural outbreaks noted that some outbreaks were not detected through surveillance of emergency department (ED) visits because symptomatic individuals did not visit an ED. Results from studies using natural and simulated outbreaks both identified the representativeness of the system as an important determinant. Systems that monitored a larger proportion of the population were more likely to detect an outbreak. Another influential system characteristic was the choice of algorithm for analyzing the data. Studies relying on simulated outbreaks suggest that for temporal surveillance, an algorithm that considers multiple days of data at each decision-point tend to outperform algorithms that consider data only from the current day. As expected, sensitivity tended to decrease as specificity increased and the time to detection also tended to increase as specificity increased.

Characteristics of the outbreak that influenced detection included the magnitude of the signal, the shape of the signal, and the timing of the outbreak. Studies using naturally occurring and simulated signals reported consistently that sensitivity increased and time to detection decreased as the magnitude of the signal increased relative to the baseline incidence. The results from simulation studies were not consistent regarding the magnitude of the signal required for consistent detection, but in general, signals with a magnitude of less than 10% of the baseline were difficult to detect when operating at high specificity. Some authors reported detecting signals with a magnitude of 30% over the baseline consistently, while other authors reported consistent detection only when the magnitude was closer to 60% over the baseline. One possible explanation for the difference in results is that signals of a given percent over baseline may be more difficult to detect with a low baseline or a high variation around the baseline.

The shape of the outbreak signal influenced outbreak detection. Simulated signals that increased in magnitude quickly over time tended to be detected more rapidly than slowly rising signals. In naturally occurring outbreaks, when cases presented over many days, sensitivity was found to be lower than when cases presented closer in time. Also, characteristics of the disease process, such as the duration of the incubation period, were shown to influence the sensitivity and timeliness of outbreak detection. These findings suggest that caution is warranted in generalizing outbreak detection results from one type of disease outbreak to another.

The timing of an outbreak influenced outbreak detection, but the influence was not consistent across studies. All but one of the simulation studies that examined the influence of timing found that outbreaks occurring on days when the baseline count was lower were detected with higher sensitivity and timeliness than outbreaks occurring on days when the baseline was higher. This result is intuitive as a signal of a set magnitude should be easier to detect when superimposed on a lower baseline as compared to a higher baseline. However, the author of one study that relied on naturally occurring disease outbreaks noted that a holiday may in practice also decrease the magnitude of a signal for an outbreak. For example, fewer people may purchase over-the-counter pharmaceuticals on a holiday and this type of behavior was not modeled in any of the simulation studies.

In terms of the methods used to evaluate outbreak detection, studies that use naturally occurring and simulated outbreaks both have strengths and weaknesses. Studies that rely on naturally occurring outbreaks can evaluate detection only for outbreaks that have occurred and these studies are usually not able to provide quantitative results. Evaluation using naturally occurring outbreaks can, however, produce rich qualitative results that help to identify practical limitations of outbreak detection and suggest topics for systematic evaluation through simulation. Simulation-based studies allow for greater flexibility and can produce quantitative results, but generalization from simulated signals to real outbreaks is not straightforward. In addition, results from evaluation studies are not reported in a consistent manner and this hampers comparison across studies.

Other authors have also noted the problem of non-standard reporting from evaluation studies of syndromic surveillance systems and have suggested adoption of a standard reporting method [7,46]. At a minimum, reports of evaluation of outbreak detection should describe the system under evaluation in terms of the characteristics included in Fig. 3. Given that evaluations using natural outbreaks tend to be qualitative, reports of these evaluations should pay particular attention to factors related to the baseline and outbreak cases and to methods used for data collection. These factors are also important for simulation studies, although for simulation studies the model of the outbreak cases should also be described. In addition, for simulation studies, authors should report the method used to combine the simulated and baseline data, the thresholds at which algorithms were evaluated and the alarm rates that correspond to the thresholds. Authors should also describe clearly the magnitude of the injected signal over time relative to the baseline incidence. Methods for reporting timeliness are still being developed, but methods that allow for simultaneous examination of sensitivity, specificity and timeliness may facilitate interpretation of the results [47].

Finally, it is important to note that the studies identified through this review do not provide evidence for many of the potential determinants of outbreak detection included in the conceptual framework presented in Fig. 3. This limitation is due, in part, to the scope of the review, but nonetheless, many unanswered questions remain. Additional fundamental research and evidence synthesis is required to understand better the influence on automated outbreak detection of outbreak and system factors in isolation and in combination.

5. Conclusions

The 35 studies identified in this review indicate that syndromic surveillance systems are capable of detecting some types of disease outbreaks rapidly with high sensitivity. The determinants of performance identified in this review include characteristics of the system and the outbreak. Influential system characteristics included the representativeness or sampling approach of the system, the outbreak detection algorithm, and the specificity of the algorithm. Important outbreak characteristics included the magnitude and shape of the signal and the timing of the outbreak.

Evaluations using natural outbreaks are best suited to answering qualitative questions, such as why certain outbreaks are not detected. Simulation of outbreaks is useful for systematic consideration of the influence of specific factors on outbreak detection and greater consistency in the reporting of evaluation results would facilitate comparison of results across studies. Future evaluations should aim to address inconsistencies in the evidence noted in this review and to identify the potential influence of other factors on outbreak detection.

Acknowledgment

Dr. Buckeridge's research is supported by a Canada Research Chair in Public Health Informatics.

References

[1] United States Government Accountability Office. Information Technology. Federal Agencies Face Challenges in Implementing Initiatives to Improve Public Health Infrastructure. 2005.

[2] 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;Suppl. 53:43–9.

[3] Loonsk JW. BioSense—A national initiative for early detection and quantification of public health emergencies. MMWR Morb Mortal Wkly Rep 2004;Suppl. 53:53–5.

[4] Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10(5):858–64.

[5] Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE II). J Urban Health 2003;80(2 Suppl. 1):i32–42.

[6] Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H, Buckeridge DL, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med 2004;140(11):910–22.

[7] 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(RR-5):1–11.

[8] National Center for Biotechnology Information. Computation of related articles, <http://www.ncbi.nlm.nih.gov/entrez/query/static/computation.html> 2003 February 6 [cited 2006 August 17].

[9] 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;Suppl 54:175–80.

[10] Buckeridge DL, Switzer P, Owens D, Siegrist D, Pavlin J, Musen M. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. MMWR Morb Mortal Wkly Rep 2005;Suppl. 54:109–15.

[11] Chen JH, Schmit K, Chang H, Herlihy E, Miller J, Smith P. Use of Medicaid prescription data for syndromic surveillance—New York. MMWR Morb Mortal Wkly Rep 2005;Suppl. 54:31–4.

[12] Das D, Metzger K, Heffernan R, Balter S, Weiss D, Mostashari F. Monitoring over-the-counter medication sales for early detection of disease outbreaks—New York City. MMWR Morb Mortal Wkly Rep 2005;Suppl 54:41–6.

[13] Edge VL, Pollari F, Lim G, Aramini J, Sockett P, Martin SW, et al. Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales. A retrospective study of waterborne outbreaks in Saskatchewan and Ontario. Can J Public Health 2004;95(6):446–50.

[14] Fleming DM, Barley MA, Chapman RS. Surveillance of the bioterrorist threat: a primary care response. Commun Dis Public Health 2004;7(1):68–72.

[15] Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci USA 2002;99(8):5237–40.

[16] Harcourt SE, Smith GE, Hollyoak V, Joseph CA, Chaloner R, Rehman Y, et al. Can calls to NHS direct be used for syndromic surveillance? Commun Dis Public Health 2001;4(3):178–82.

[17] Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 2004;10(5):858–64.

[18] Hogan WR, Tsui FC, Ivanov O, Gesteland PH, Grannis S, Overhage JM, et al. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. J Am Med Inform Assoc 2003;10(6):555–62.

[19] Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. Ann Emerg Med 2003;41(4):447–52.

[20] Ivanov O, Gesteland PH, Hogan W, Mundorff MB, Wagner MM. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. AMIA Annu Symp Proc 2003:318–22.

[21] Kleinman KP, Abrams A, Mandl K, Platt R. Simulation for assessing statistical methods of biologic terrorism surveillance. MMWR Morb Mortal Wkly Rep 2005;Suppl. 54:101–8.

[22] Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space–time permutation scan statistic for disease outbreak detection. PLoS Med 2005;2(3):e59.

[23] Marx MA, Rodriguez CV, Greenko J, Das D, Heffernan R, Karpati AM, et al. Diarrheal illness detected through syndromic surveillance after a massive power outage: New York city, August 2003. Am J Public Health 2006;96(3):547–53.

[24] Miller B, Kassenborg H, Dunsmuir W, Griffith J, Hadidi M, Nordin JD, et al. Syndromic surveillance for influenza-like illness in ambulatory care network. Emerg Infect Dis 2004;10(10):1806–11.

[25] Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for communitywide influenza-like illness, New York City. J Urban Health 2003;80(2 Suppl. 1):i43–9.

[26] Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg Infect Dis 2003;9(6):641–6.

[27] Nordin JD, Goodman MJ, Kulldorff M, Ritzwoller DP, Abrams AM, Kleinman K, et al. Simulated anthrax attacks and syndromic surveillance. Emerg Infect Dis 2005;11(9):1396–400.

[28] Ozonoff A, Forsberg L, Bonetti M, Pagano M. Bivariate method for spatio-temporal syndromic surveillance. MMWR Morb Mortal Wkly Rep 2004;Suppl. 53:59–66.

[29] Quenel P, Dab W. Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. Eur J Epidemiol 1998;14(3):275–85.

[30] Reis BY, Mandl KD. Integrating syndromic surveillance data across multiple locations: effects on outbreak detection performance. AMIA Annu Symp Proc 2003:549–53.

[31] Reis BY, Mandl KD. Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. Ann Emerg Med 2004;44(3):235–41.

[32] Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2.

[33] Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci USA 2003;100(4):1961–5.

[34] 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;Suppl. 54:151–6.

[35] Rodman JS, Frost F, Jakubowski W. Using nurse hot line calls for disease surveillance. Emerg Infect Dis 1998;4(2):329–32.

[36] Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. MMWR Morb Mortal Wkly Rep 2004;Suppl 53:152–8.

[37] Stoto MA, Schonlau M, Mariano LT. Syndromic surveillance: Is it worth the effort? Chance 2004;17(1):19–24.

[38] Tsui FC, Wagner MM, Dato V, Chang CC. Value of ICD-9 coded chief complaints for detection of epidemics. Proc AMIA Symp 2001:711–5.

[39] Wallstrom GL, Wagner M, Hogan W. High-fidelity injection detectability experiments: a tool for evaluating syndromic surveillance systems. MMWR Morb Mortal Wkly Rep 2005;Suppl. 54:85–91.

[40] Wang L, Ramoni MF, Mandl KD, Sebastiani P. Factors affecting automated syndromic surveillance. Artif Intell Med 2005;34(3):269–78.

[41] Weber SG, Pitrak D. Accuracy of a local surveillance system for early detection of emerging infectious disease. JAMA 2003;290(5):596–8.

[42] Yih KW, 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;Suppl. 54:157–62.

[43] Buehler JW. Surveillance. In: Rothman KJ, Greenland S, editors. Modern epidemiology. Philadelphia, PA: Lippincott-Raven; 1998. p. 435–57.

[44] Dato V, Wagner MM, Fapohunda A. How outbreaks of infectious disease are detected: a review of surveillance systems and outbreaks. Public Health Rep 2004;119(5):464–71.

[45] Hopkins RS. Design and operation of state and local infectious disease surveillance systems. J Public Health Manag Pract 2005;11(3):184–90.

[46] Sosin DM, DeThomasis J. Evaluation challenges for syndromic surveillance–making incremental progress. MMWR Morb Mortal Wkly Rep 2004;Suppl. 53:125–9.

[47] Kleinman K, Abrams A. Metrics for assessing the performance of spatial surveillance. Statist Methods Med Res 2006;15(5).